**Fake News Detection Using Deep Learning Models**

**Background**

Fake news is fictitious information in the form of articles, stories or hoaxes that are deliberately false and are created to deliberately misinform or deceive (Monther and Ali 2018). Fake news has taken many different forms across history, ranging from word of mouth, printed form, to the internet we have today. Before the internet, most people got their information from a few trusted sources, such as newspapers or television networks, so it was in the best interest of these sources to avoid distributing fake news to maintain credibility. Ever since the internet, we have been bombarded with so much information form so many sources, the internet has made it so much easier to spread fake news. Studies show that 75% of people who see fake news think its real news (Craig and Singer-Vine 2016). Having disinformation online can have a huge impact on our society and our democracy (Colomina 2021).

The findings of a study done in 2006 supports the idea that the more time people spent on a medium, the more credibility they gave to it (Althaus 2006). Since most people spend more time on and get news from social media, we’ll be focusing on social media for this study.

**Aim**

This project seeks to build a deep learning model that can accurately detect fake news in social media content. We will provide a practical solution for this problem in the form of an interactive web application.

**Key Techniques:** Deep learning, Transfer Learning, Text Preprocessing, Fine-tuning.

**Objectives**

1. To provide a real world solution for fake news detection on social media.
2. To preprocess dataset using tools like transformers library from hugging face as well as helper functions.
3. To build a traditional deep learning model and use it as our baseline model.
4. To train non-transformer based models and compare results using metrics like f1 score and recall to evaluate performance.
5. To combine datasets based on articles that cover a wide range of topics including politics, all sourced from social media.
6. To host the best model on Hugging face so that it may be available publicly.
7. To deploy our chosen model as an interactive web app using an API or python libraries like Streamlit & Gradio. The web application will accept a URL as input and classify the content of the URL as either fake or reliable.

**Methodology**

The process we will use for this research project is summarized in the following steps:

1. Data Selection: A dataset or datasets of fake and real news articles scraped from social media platforms will be chosen. The dataset will be split into a test set and a train set.
2. Preprocessing: Next, we will prepare the dataset for modeling. There are three necessary steps involved:

* Data Cleaning: using helper functions to remove punctuations, usernames, links, etc. as well as formatting to lowercase.
* Encoding: We must encode the labels using label encoder or using another helper function.
* Tokenization: we will use a tokenizer to convert the text into numerical tokens.

1. Training: We will first train a deep learning model like Support Vector Machines or Naive Bayes as our baseline model. Then we will finetune at least three pretrained transformer-based models from hugging face with just 10 epochs.
2. Evaluation: We will use f1 score as the metric to evaluate the performance of our models.
3. Deployment: The best model will be deployed as a web application using Streamlit.

**Literature Review**

Hei et al. (2022) in his study uses a unified training approach by using a pretrained BERT model, using a combination of 3 datasets including the ISOT fake news dataset & 2 datasets from Kaggle. The proposed unified training strategy achieved an F1 score of 0.97 with a 97% accuracy, which surpassed the existing models like Random Forests, Convolutional Neural Networks (CNN), Long Short Term Memory (LSTM), etc.

Deep learning models generally perform better than machine learning models due to the increasing data samples, but this typically requires a larger dataset and increases training time significantly. This study solves this problem by fine-tuning a model with prior knowledge of the task, therefore the model does not require a very large dataset since it is not built from scratch. Moreover, there was a significant reduction in training time by almost 1.5x after removing words with less than 3 letters from the combined input dataset. This study has two major limitations:

1. The use of only the BERT model whereas there are other models that could achieve better results.
2. The study failed to account for the impact of social media which plays a major role in spreading fake news.

In a study published by Joshi et al. (2021), a dataset of covid-19 related news articles was used to train various deep learning models like CNN, LSTM, HAN, etc. as well as transformer-based models like BERT and DISTILBERT. Using accuracy score to evaluate, the transformer-based models, both pretrained and fine-tuned outperform the other models with an absolute difference of 3-4%. The best performing model was BERT with an accuracy of 98.41%.

However, this study has some limitations. My major concerns are listed below:

1. The dataset used is based on news articles related to covid-19. So, the results of the study may not be applicable to other topics.
2. The study uses only accuracy as its evaluation metric. Other metrics such as precision and recall could be more informative.

A more recent study done in 2022 by Althabiti et al. (2022) indicates that BERT transformer-based models perform better than other models. Traditional machine learning models were compared with BERT based pretrained models using data from 2010 to 2022 that is associated with different topics. An external dataset called the FakeNews Classification dataset was added to the dataset for the experiment, but this did not improve performance. This study did not have a social media focus and used only f1 score and accuracy to evaluate performance.

For this study, we’ll be using dataset(s) sourced from social media to train our models. We will also explore other evaluation metrics apart from accuracy as well.

**Summary**

When fake news affects public opinion, it could sow discord and incite violence. It even has the power to change our democracy. It is our duty to do everything in our power to stop the spread of fake news. This project aims to use the deep learning techniques to build a model that can classify news as either fake or not fake, therefore providing the public with a web application that can detect fake news in social media content.

**References**

Xuewen Zhang, 2019. Deep learning methods for Fake News detection.

Link: https://rb.gy/1i1i4

Monti et al., 2021. Unified Fake News Detection using Transfer Learning of Bidirectional Encoder Representation from Transformers model.

Link: https://rb.gy/59sl6

Joshi et al., 2021. Evaluating Deep Learning Approaches for Covid19 Fake News Detection.

Link: https://arxiv.org/pdf/2101.04012

Aldwairi and Alwahedi, 2018. Detecting Fake News in Social Media Networks.

Link: https://www.sciencedirect.com/science/article/pii/S1877050918318210

Craig and Singer-Vine, 2016. Most Americans who see fake news believe it; new survey says.

Link: https://www.buzzfeednews.com/article/craigsilverman/fake-news-survey

Colomina, 2021. The impact of disinformation on democratic processes and human rights in the world.

Link: https://www.europarl.europa.eu/RegData/etudes/STUD/2021/653635/EXPO\_STU(2021)653635\_EN.pdf