Fake Information Prediction in Social Media Using Deep Learning

Rohan Bera

Georgia State University

Atlanta, GA, 30303

rbera1@student.gsu.edu

Sai Roopesh Mandava

Georgia State University

Atlanta, GA, 30303

smandava4@student.gsu.edu

Meghana Puli

Georgia State University

Atlanta, GA, 30303

mpuli1@student.gsu.edu

Abstract

Online sources of information have become a massive part of the smart society. Blog posts, social media and other online platforms have made obtaining news quick and easy. While this has improved the spread of knowledge, this has also led to the widespread production of false news. With a large variety of online data, fact-checking every piece of information is a mammoth task. So in order to curb such counterfeit news from being spread, we need some tools to automate the process and find efficient ways to classify it. Our aim is to find a reliable and accurate model that classifies a given news article as either fake or authentic using deep learning and LSTM.

1 Project title

Through our project, **Fake Information Prediction**, we aim to build a model which classifies whether the news given is genuine news or false news to avoid panic among the people and bring awareness. This project undergoes text analytics, an artificial intelligence methodology that uses natural language processing (NLP) and deep learning. This goal is to be reliable and produce more accurate results.

2 Dataset

- Word Embedding over Linguistic Features for Fake News Detection (WELFake) dataset [2] consists of 72,134 news articles, with 35,028 real and 37,106 fake news. The dataset contains four columns: Serial number (starting from 0); Title (about the text news heading); text (about the news content); and Label (0 = fake and 1 = real).

- Fake and real news dataset [1] contains two files. One file contains all the fake news, and the other file contains verified real news. Each file contains eight columns, including the serial number, title, article, subject etc.

3 Project idea description

In recent times, social media has emerged as the primary source of information for young individuals. Given its widespread usage and convenience, online content in the form of posts, articles, and blogs has become the most suitable medium for obtaining knowledge and staying informed about current events. However, unlike traditional forms of news delivery, such as newspapers and news channels, where trained editors and reporters fact-check the information before publishing, social media platforms lack a robust verification process. This is because the content posted on social media is generated by the users themselves, without any form of scrutiny or editing.

This project aims to build a model that classifies whether the news is genuine or false to avoid panic among the people and bring awareness. We aim to do this with the help of deep learning.

4 Teammates and milestones

4.1 Teammates

We have divided the total project into three broad categories:

- **Literature review**

  We researched several research papers and articles and pooled our knowledge. After lots of discussions, we selected the topic *Fake information detection*. We listed out research articles that best suit our project based on their proposed algorithm and methodology.

- **Finding appropriate Datasets**

  Finding appropriate datasets is of utmost importance for training our dataset. We found two datasets that were a good match to our use case. The datasets can be found in the section above.

- **Training the model**

  We will be training the model using LSTM. We will adjust the hyper-parameters as we progress with the project.

We divided each topic among each of us equally. We researched several research papers and articles and pooled our knowledge. After lots of discussions, we selected the topic **Fake information detection**. We listed out research articles that best suit our project based on their proposed algorithm and methodology.

4.2 Milestones

We have divided the project into tasks and assigned a fixed number of days to each of them. The timeline can be seen in Table 1.

Days	Tasks
1	Researching required libraries
1	Finding appropriate datasets
1	Pre-processing of data
1	Vectorization
7	Implementing Algorithms
2	Observations and results
2	Documentation

Table 1: Group mile-stones.

5 Software

We will be using **Python** to develop our model. Within python, we plan on using the following libraries within our project:

- **NLTK**
  NLTK is a complete toolkit for all NLP techniques. We will be using this to perform deep learning algorithms on our data.
- **Pandas**
  We will be using pandas to visualise large datasets in human-friendly table structures.
- **Matplotlib**
  We will be using matplotlib to plot any necessary graphs that will help us understand important trends. Using this library, we will also be able to visualise the patterns being followed by the training set and test set errors.

Finally, we will be using **Google Colab**, or "Colaboratory" to execute our algorithm. Colab allows users to write and execute Python in their browser, with zero configuration, access to GPUs free of charge and easy sharing. The entire project will be open source and will be hosted on **GitHub**.

References

[1] Hadeer Ahmed, Issa Traore, and Sherif Saad. Detecting opinion spams and fake news using text classification. *Security and Privacy*, 1:e9, 12 2017.

[2] Pawan Verma, Prateek Agrawal, Vishu Madaan, and Radu Prodan. Mcred: multi-modal message credibility for fake news detection using bert and cnn. *Journal of Ambient Intelligence and Humanized Computing*, pages 1–13, 07 2022.