Detection of fake news in the medical field

1st Semester of 2023-2024

First Author
rares.tudur@s.unibuc.ro

Second Author
darius-stefan.corneciu@s.unibuc.ro

Abstract

Knowing the impact that **fake news** has on public health, this project aims to develop a system to identify and classify fake news in the medical field. **The objective** is to detect various *linguistic patterns* specific to fake news in the medical domain by *collecting and processing articles* that will be used to train a model.
1, 3, 4, and References.

1 Introduction

Fake news has been present in the world since the invention of *Gutenberg's printing press*. However, identifying the first case in the medical field is challenging. A well-known case in America dates back to 2017, published by a site called **eternally.com**. The article was titled *"Dandelion weed can boost your immune system and cure cancer"*, a claim that lacked any scientific basis. In **Romania**, the spread of fake news in the medical field peaked during the pandemic, with the dissemination of information aimed at *questioning vaccines and public health measures*.
In the following list, we will present the contribution of each team member:

- **Tudur Rares**: Research for the BERT model
- **Corneciu Darius**: Research in the field of fake news

A summary of the approach involves training a model on a dataset from a database. After completing the training, the system will present the user with a text box where they can input the title of an article, and the model will determine whether it is fake news or not.
Through this project, we aim to identify headlines with specific patterns and mark them as fake news, helping the population stay correctly informed and preventing a *"disinformation pandemic"* from resurfacing.

We conducted research to understand the origins of fake news and its historical uses. On the implementation side, we explored similar models and methods to understand how various models work, enabling us to select the one that best fits our requirements.
Here are a few papers related to fake news in the medical field:

- (Hou et al., 2019) They address the issue of misinformation in online videos, focusing on videos about prostate cancer. The authors manually selected 250 labeled videos. The proposed models achieved an accuracy of up to 74

- (Vladika et al., 2023) The paper proposes a new dataset, HealthFC, which contains 740 health-related statements in both German and English. These statements are labeled for veracity by domain experts. The authors, using Natural Language Processing (NLP), suggest that the dataset can be utilized for statement verification, evidence retrieval, and generating explanations.

- (D et al., 2021) This paper proposes a graph-based social framework and attention mechanism for the early detection of fake news in the health domain. The study demonstrates the effectiveness of the approach in capturing fake news in the medical field at early stages.

- (Srba et al., 2022) The paper presents an extensive dataset of approximately 317,000 medical articles and blogs, intended for the study of medical misinformation.

In this **project**, we learned how fake news works and the *linguistic elements* to look for in order **to avoid such news**. In the future, *we aim to continue studying* this field, as it is one that will **never end**, given that some ***people always tend to believe something at first glance***.

Library	Used for
Torch	*A deep learning framework used for building and training neural networks.*
Transformers	*A library by Hugging Face that provides pre-trained models for natural language processing (NLP) tasks, including BERT, GPT, and other transformer-based models.*
Datasets	*A library from Hugging Face that provides easy access to a wide range of datasets.*
Scikit-learn	*A popular Python library for machine learning that provides simple and efficient tools for data mining and data analysis, including algorithms for classification, regression, and clustering.*

Table 1: Libraries and their usage in the project.

2 Approach

In this section, we will present detailed aspects of our approach regarding our project. Initially, we worked separately, and the final version of the code can be found at the following GitHub link.
As software, I used Python along with several libraries, which are presented in Table 1.

The training time varied depending on the different computers used. The approximate time is 5 minutes for the dataset present in *data.csv*, which contains a total of 10,000 titles. The dataset was taken from **Kaggle**.

In the code, we are using a deep learning approach for **binary classification of news headlines** *(real vs fake)*. The specific tools and architecture used are:

- **Hugging Face Transformers**: This library is used to load and fine-tune transformer models such as *BERT*.
- **PyTorch**: It is used for defining the model and setting up training.
- **BERT Model**: Specifically, *AutoModelForSequenceClassification* is used, which is pre-trained on a large corpus and fine-tuned for the task at hand
- **BertTokenizer**: The tokenizer splits the input text into subwords and converts them into numerical tokens.
- **Trainer API**: From Hugging Face, it is used to manage the training and evaluation process with minimal code.

The evaluation report is generated using **classification report** from *sklearn.metrics*. This provides the following key metrics:

- **Precision**: The ratio of correctly predicted positive observations to the total predicted positives.
- **Recall**: It answers the question: "Of all the actual real news, how many were predicted as real?"
- **F1-score**: The weighted average of Precision and Recall(*false positives and false negatives*)
- **Support**: The number of actual occurrences of the class in the dataset

3 Limitations

3.1 Language Limitations

The model, uses the *BERT* model pretrained on the "*bert-base-uncased*" architecture, which is primarily designed for *English text*. Although BERT has multilingual versions (*bert-base-multilingual-uncased*), this model does not support other languages by default.

3.2 Scalability Issues

BERT models, while powerful, are *computationally intensive*. **Fine-tuning** BERT for large datasets or longer texts can be *resource-heavy* and longer training times.

3.3 GPU Resource Requirements

The model currently runs on a GPU if available (*torch.cuda.isavailable()*), but BERT models typically **require high memory and processing power for training**. This could be a limitation for anyone without access to high-performance hardware.

3.4 Interpretability and Explainability

Transformer models, including **BERT**, are often seen as *"black-box"* models, making it difficult to explain how they arrive at their predictions.

4 Conclusions and Future Work

Now that we have completed the project, we would probably have opted for a different model. While BERT is a powerful model, for our small dataset, it might be overkill. We could have explored smaller models like DistilBERT.

We could have improved this project using techniques like gradient clipping, batch normalization, or experimenting with learning rate schedules and we could have achieved greater improvement if we had a larger dataset.

Honestly, the project was interesting, and we learned new things about artificial intelligence models, though they weren't to our liking. At the same time, we learned how to use Overleaf for document creation and gained a deeper understanding of fake news.

References

Prasannakumaran D, Harish Srinivasan, Sowmiya Sree S, Sri Gayathri Devi I, Saikrishnan S, and Vineeth Vijayaraghavan. 2021. *SOMPS-Net: Attention-based Social Graph Framework for Early Detection of Fake Health News*.

Rui Hou, Verónica Pérez-Rosas, Stacy Loeb, and Rada Mihalcea. 2019. *Towards Automatic Detection of Misinformation in Online Medical Videos*.

Ivan Srba, Branislav Pecher, Matus Tomlein, Robert Moro, Elena Stefancova, Jakub Simko, and Maria Bielikova. 2022. *Monant Medical Misinformation Dataset: Mapping Articles to Fact-Checked Claims*.

Juraj Vladika, Phillip Schneider, and Florian Matthes. 2023. *HealthFC: Verifying Health Claims with Evidence-Based Medical Fact-Checking*.