FAKE NEWS DETECTOR

Anuj Timsina, Prajwal Chhetri, Sachi Kamat, Sandip Thapa

Department of Computer and Electronics Engineering, Purwanchal Campus, IoE, TU, Dharan

Background:

False information has proven to be a serious issue in the present context due to rise in use of social media. It is now extremely easy to spread false information because people on social media are not responsible enough to check the authenticity of a news before sharing it.

To validate the integrity of these data, we explore the application of Natural Language Processing techniques to identify when a news source may be producing a fake news.

The model focused on identifying fake news sources based on multiple articles originating from a source. Once a source is labeled as a producer of fake news, we can predict with high confidence that any future articles from that source will also be fake news

Introduction

Traditionally we got our news from trusted sources, journalists and media outlets that are required to follow strict codes of practice. However, the internet has enabled a whole new way to publish, share and consume information and news with very little regulation or editorial standards. Hence, we have observed an increase in spread of false information. Public is entitled to authentic news and information. This project is dedicated to ensure the authenticity of news.

General objectives:

- To check authenticity of news from various sources using Natural Language Processing (NLP).

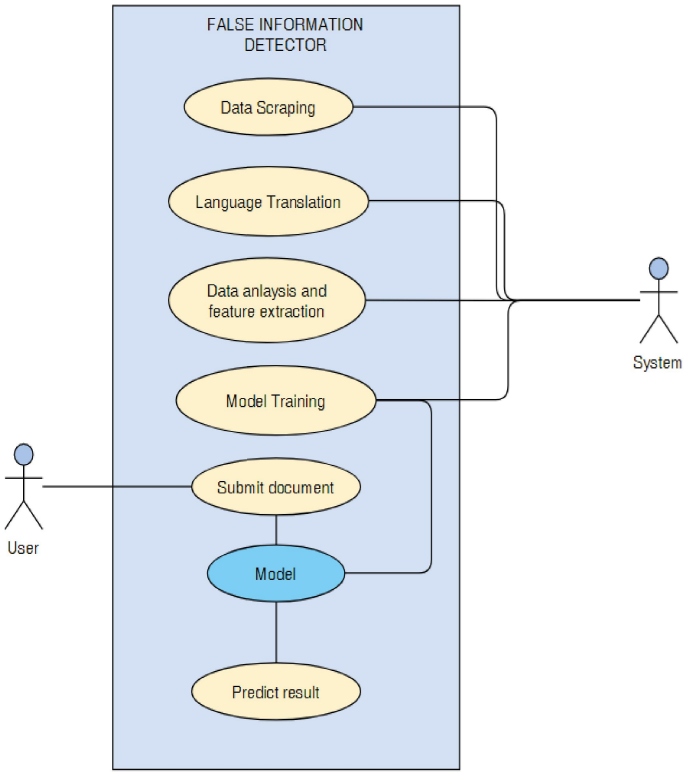
Specific objectives:

- To scrape data from various sources using BeautifulSoup.
- To analyse and extract features of the collected data samples using Python libraries like Keras, Tensor Flow, NumPy, Pandas, Matplotlib, SciKit-Learn.
- To make use of NLP model training. To predict whether the user input information is authentic or not using the machine learning model.

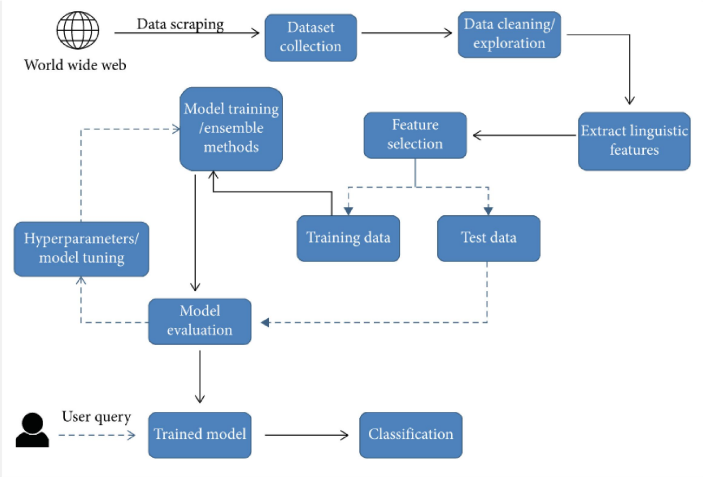
Features:

I. To identify fake news.
II. To provide authentic source of information.
III. To show the application of NLP in real world.

Use Case Diagram



Methodology:



For our project development the following aspects are used:

i. NLP

NLP uses supervised machine learning technique. It goes through the following steps in order to train NLP model.

1. Tokenization: breaks down text into smaller semantic units or single clauses
2. Part-of-speech-tagging: marking up words as nouns, verbs, adjectives, adverbs, pronouns, etc.
3. Stemming and lemmatization: standardizing words by reducing them to their root forms
4. Stop word removal: filtering out common words that add little or no unique information, for example, prepositions and articles (at, to, a, the).

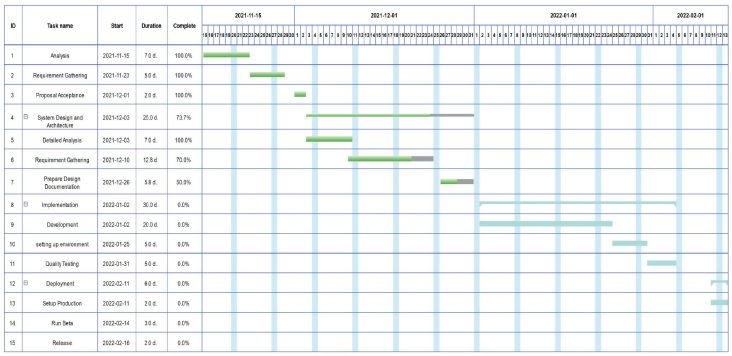
ii. Development Platform

We are developing our project using JavaScript programming language and Python for model training and predicting the result. We are using Node.js, Express.js for back-end development and React.js for front-end development. Since our team has 4 members, we are developing our project using GitHub which makes it easier to contribute in the project. Also, the IDE we are using is Visual Studio code.

Expected Outcome:

- To address the authenticity of growing online news of certain sources and verify the sources.
- To create a model which is expandable for further data sets in the future.

Gantt Chart:



References:

- www.youtube.com
- www.wikipedia.org
- https://www.hindawi.com/journals/complexity/2020/8885861/
- https://www.ibm.com/cloud/learn/natural-language-processing
- Indurkhya, N., & Damerau, F. J. (Eds.). (2010). Handbook of natural language processing (Vol. 2). CRC Press.
- Müller, A. C., & Guido, S. (2016). Introduction to machine learning with Python: a guide for data scientists. " O'Reilly Media, Inc.".
- Fake News Detection with Machine Learning, using Python. Medium. (2021). Retrieved 29 December 2021, from https://towardsdatascience.com/fake-news-detection-with-machine-learning-using-python-3347d9899ad1.wi.com/ journals/complexity/2020/8885861/fig1/.