**Artificial Intelligence Internship**

**Project Report**

Project Title:

**FAKE NEWS ANALYSIS**



Team Members:

Pranay Thipirishetty – 19BAI1056

Mahalaxmi Bashetty – 19BAI1004

Kandagatla Uttej Kumar -19BAI1015

Potti Sai Pavan Guru Jayanth – 19BAI1045

**Index:**

|  |  |  |
| --- | --- | --- |
| ***S.No*** | ***Format*** | ***Page-No*** |
| 1 | Introduction | 3 |
| 2 | Problem Statement | 3 |
| 3 | Solution | 4 |
| 4 | Literature Survey | 5 |
| 5 | Experimental investigation | 6-7 |
| 6 | Software specifications | 8 |
| 7 | Flow chart | 9 |
| 8 | Results | 10-11 |
| 9 | Conclusion | 12 |
| 10 | Future scope | 13 |
| 11 | Bibliography | 13 |

**INTRODUCTION:**

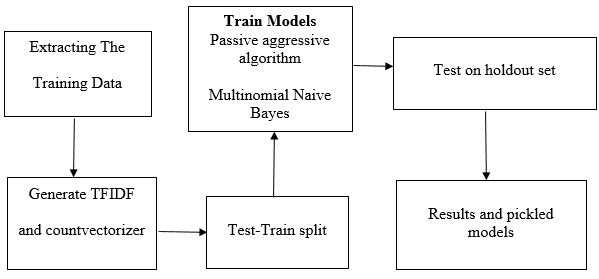
Fake news and hoaxes have always existed, even before the Internet. Fake news on the internet is defined as “fictional content created with the intent of deceiving readers.” Fake news is published on social media and news sites to promote readership or as psychological warfare. In general, the objective is to benefit on clickbait's. Clickbait's utilise flashy headlines or graphics to encourage visitors to click links to increase ad income. This study shows that the prevalence of fake news stories considering the advancements in communication made possible by the rise of social networking sites. The aim of the project is to provide a system that consumers may use to detect and filter out sites that contain incorrect and misleading information. To accurately identify fraudulent postings, we employ basic and carefully chosen elements of the title and content.

Problem Statement:

The project's goal is to find a solution that can be used to recognise and filter out sites that contain whether it is real news or fake news, therefore assisting consumers in avoiding being enticed by clickbait. It is critical that such solutions be discovered since they will be beneficial to both readers and the IT firms engaged in the issue.

Solution:

In this project, a model based on the count vectorizer or a tfidf matrix (i.e., word tallies related to how frequently they are used in other articles in our dataset) might be effective. Because this is a text classification problem, it is recommended to use a Naive Bayes classifier, which is typical for text-based processing. The primary plan is to create a model for text transformation (count vectorizer versus tfidf vectorizer) and decide which type of text to utilise (headlines vs full text). The next step is to extract the most optimal features for count vectorizer or tfidf-vectorizer, which is done by using an n-number of the most used words and/or phrases, lower casing or not, removing stop words such as “the,” “when,” and “there,” and only using words that appear at least a given number of times in each text dataset.



Literature survey:

Many automatic detection approaches for fake news and deception posts have been reported in the literature. Because there are many different components to fake news detection, from employing chatbots to disseminate disinformation to utilising clickbait's to propagate rumours. There are several clickbait's accessible on social media networks such as Facebook, which increase the sharing and like of content, therefore spreading false information. A great deal of effort has gone into detecting false information.

The Author [1] discussed Linguistic Cue Methods with Machine Learning, Bag of Words Approach, Rhetorical Structure and Discourse Analysis, Network Analysis Approaches, and SVM Classifiers. These algorithms are only text-based and offer little or no improvement over previous approaches.

The Author [2] examined the concepts, methodologies, and algorithms used for classifying false and manufactured news articles, authors, and subjects from virtual communities, as well as assessing their scope and efficiency. The report also proposed research challenges based on previously unknown aspects of false news and varied relationships between news pieces, writers, and subjects. The authors of the study discuss Fake Detector, an automated fake news inference algorithm. It is based on textual classification and employs a deep diffusive network model to concurrently learn these same portrayals of news items, authors, and subjects. Fake Detector focuses on two key elements: The wide diffusional proposed approach will be composed of representations features extraction and reliability labelling inference. Fake Detector.

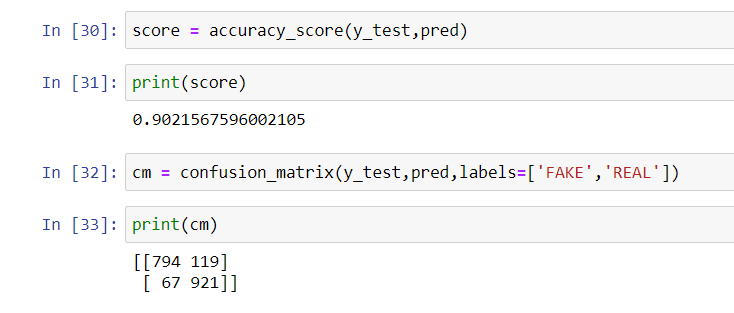
Experimental investigations:

In this project, a model based on the count vectorizer or a tfidf matrix (i.e., word tallies related to how frequently they are used in other articles in our dataset) might be effective.

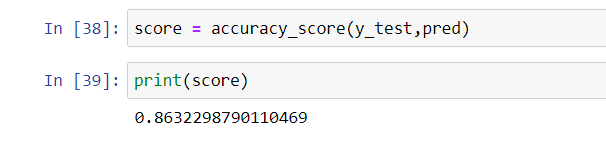
The primary plan is to create a model for text transformation (count vectorizer versus tfidf vectorizer) and decide which type of text to utilise

Here are the experimental results of our primary plan which is count vectorizer versus Tfidf Vectorizer

* Tfidf Vectorizer Accuracy

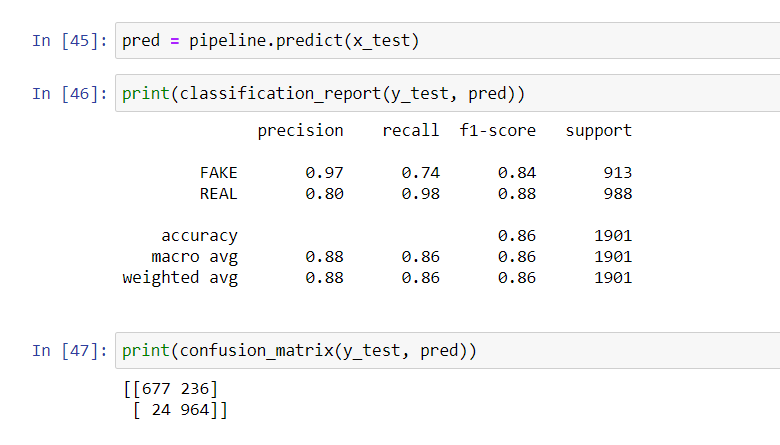


* Count Vectorizer Accuracy



From the above figures we can conclude that the Tfidf Vectorizer Accuracy is more. So, we will be using Tfidf Vectorizer algorithm to train the model.

* Classification report



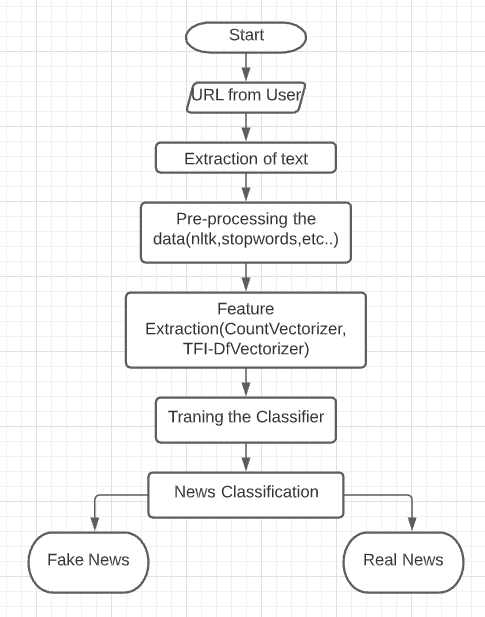
Software specifications:

* flask-Cors
* NumPy
* newspaper3k
* Sklearn
* pandas
* flask
* Nltk

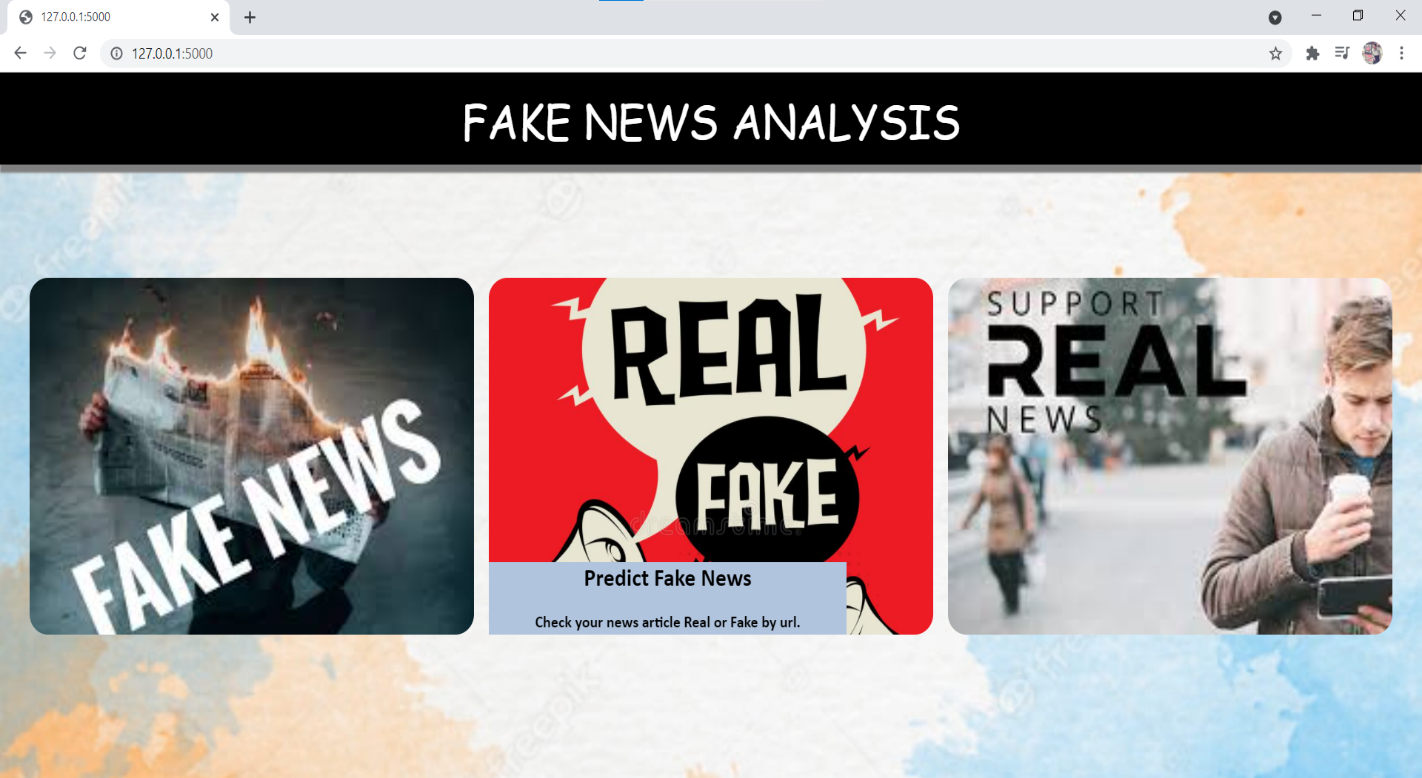
In nltk :

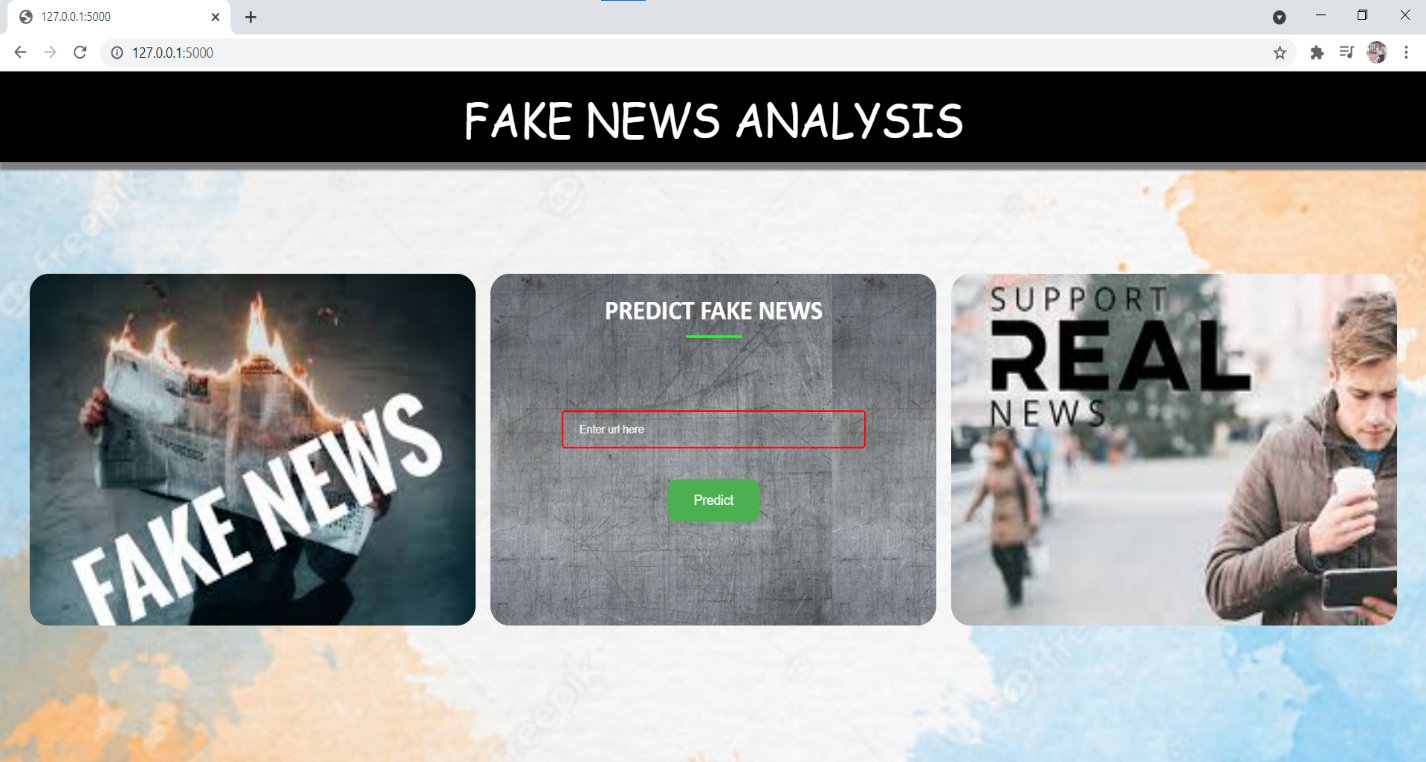
* Brown
* Punkt
* Maxent\_Treebank\_pos\_tagger
* Movie\_Reviews
* wordnet
* Stop words

Flow chart:

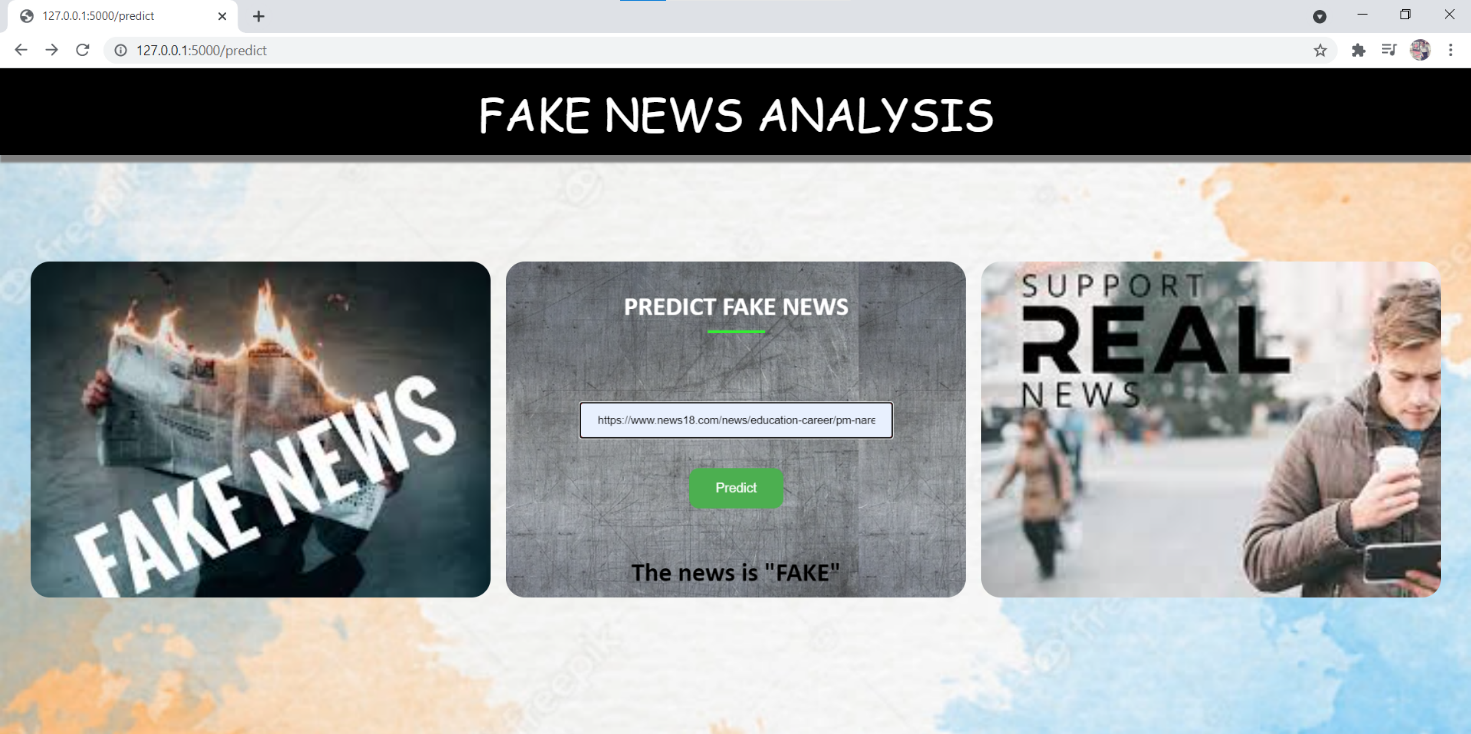


Results:

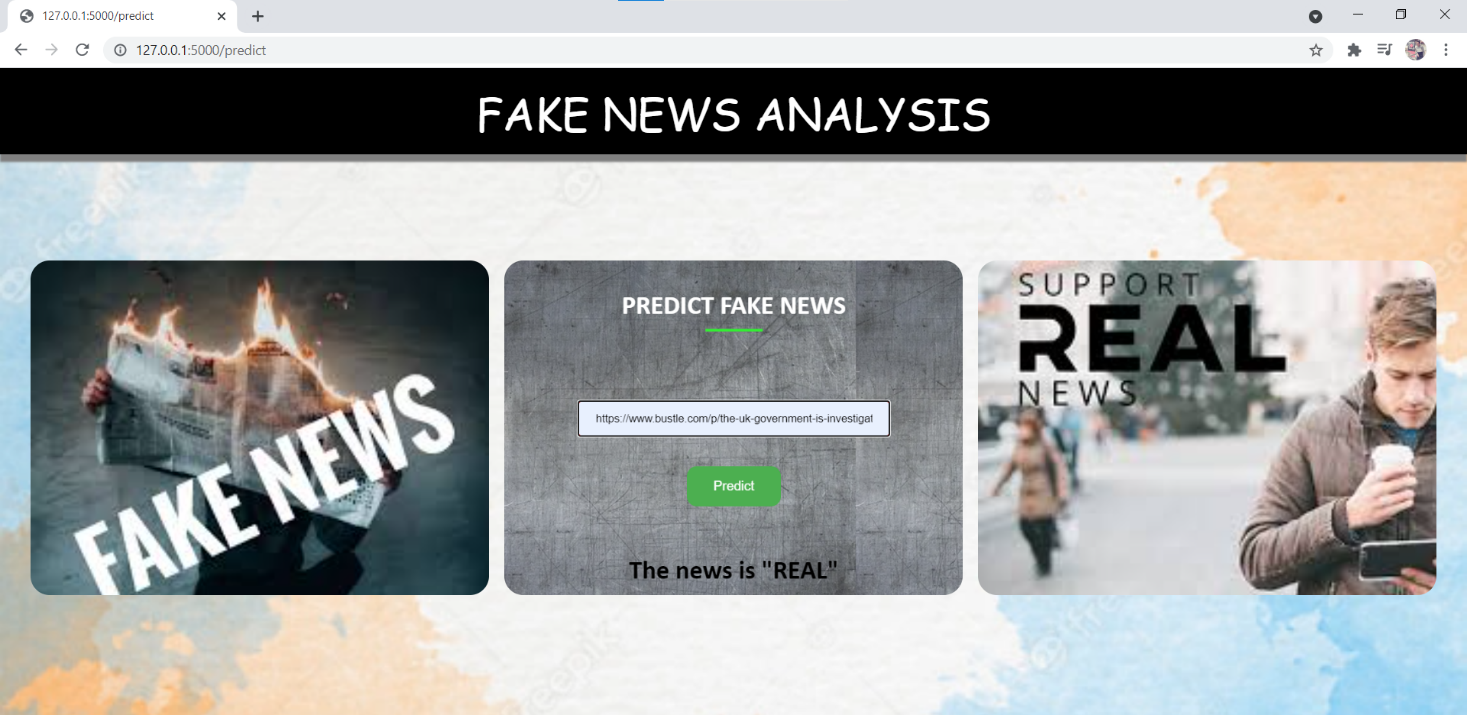




Predicting Fake news:



Predicting Real news:



Conclusion:

The bulk of tasks in the twenty-first century are completed online. Newspapers, which were formerly favoured in physical copy, are increasingly being replaced by internet apps such as Facebook, Twitter, and news articles. The rising problem of fake news only complicates matters and attempts to alter or sway people's opinions and attitudes regarding the usage of digital technologies. When a person is duped by fake news, one of two things might happen. People begin to believe that their assumptions about a certain issue are correct. Another issue is that even if there is a new article accessible that contradicts an allegedly false one, people trust in words that simply reinforce their beliefs without considering the facts involved.

As a result, in order to combat the problem, Google and Facebook are taking efforts to prevent the spread of fake news. Our algorithms categorise information from a URL as real or fake. NLP and Machine Learning Techniques are utilised to achieve this. Any user with a browser and an internet connection can use the website to verify the validity of the new article.

Future scope:

From this project we can also extend this project in future by employment of a technology that can identify and eliminate fraudulent sites from the results presented to a user by a search engine or a social media news feed is advocated as a solution to the issue of false news. The user can download the tool and then add it to the browser or programme they use to get news feeds. Once operational, the tool will use various techniques including those related to the syntactic features of a link to determine whether the same should be included as part of the search results.

Bibliography:

* <https://www.ijariit.com/manuscripts/v4i1/V4I1-1432.pdf>
* <https://arxiv.org/ftp/arxiv/papers/2102/2102.04458.pdf>
* <https://medium.com/swlh/fake-news-detection-using-machine-learning-69ff9050351f>
* [1]. Conroy, N. J., Rubin, V. L., & Chen, Y. (2015, November). Automatic deception detection: Methods for finding fake news. In Proceedings of the 78th ASIS&T Annual Meeting: Information Science with Impact: Research in and for the Community (p. 82). American Society for Information Science.
* [2] Zhang, J., Cui, L., Fu, Y., & Gouza, F. B. (2018). Fake news detection with deep diffusive network model. arXiv preprint arXiv:1805.08751.