Fake News Classification Using NLP

Ms.G .Hemasudharani[1], G.Nikhitha[2], N S Praneetha Gandikota[3], S. Sri Divijendra Kumar[4],Md.Abdul Naveed[5]

[1]Assistant Professor, Department of CSE-Artificial Intelligence and Machine Learning , S.R.K Institute ofTechnology, NTR, Andhra Pradesh, India.
[2]*student, Department of CSE-Artificial Intelligence and Machine Learning, S.R.K Institute ofTechnology, NTR, Andhra Pradesh, India*
[3]*student, Department of CSE- CSE-Artificial Intelligence and Machine Learning, S.R.K Institute of Technology, NTR, Andhra Pradesh, India*
[4]*student, Department of CSE- Artificial Intelligence and Machine Learning, S.R.K Institute ofTechnology, NTR, Andhra Pradesh, India*
[5]*student, Department of CSE- Artificial Intelligence and Machine Learning, S.R.K Institute ofTechnology, NTR, Andhra Pradesh, India*

Abstract— In today's modern world, "fake news" has been a major concern, spreading like wildfire through many platforms. This phenomenon not only undermines the credibility of information but also misleads society. Nowadays, social media is the greatest means by which fake news spreads all over the place. This can cause many problems such as defamation of people and spreading news in favour of specific individuals. Fake news often targets the most prominent, powerful, and influential people in society, aiming to tarnish their reputation. The escalating impact of fake news knows no bounds. Fake news is often biased, favouring a single person or a section of people in society for their personal benefits. To mitigate these challenges and promote transparency, there is a need to reduce the spread of fake news. Introducing a "Fake News Classifier using NLP" offers a promising solution to combat this issue. By using machine learning algorithms, this classifier can effectively identify misleading information as fake news, thereby contributing to awareness in society and reducing losses.

Keywords— Natural Processing Language, TF-IDF, Flask, Classification, MultinomialNB, Accuracy.

1. INTRODUCTION

Fake news primarily consists of mis leading information spread across the society, creating turmoil. In this era, Information is all over and the number of people accessing the information is increasing substantially. There should be awareness among users regarding what type of information they are consuming - "is it real? or fake?". Moreover, most of the social media platforms allow users to share their views through stories, statuses, posts, directly affecting the spread of news, which may often considered fake. One very famous Social media platform, what's App serves as a means for consistently sharing fake news among its users through What's App groups, Statuses, personal messages. If this sharing or spreading of fake news reaches a significant number, there is a risk of people believing it, leading to disorder.

One such recent example is the rumour of the ban on 10 rupees coin in India. There was widespread news that 10 rupees coins in India were banned, thanks to social media, which facilitated the rapid spread of this misinformation. Nobody was accepting 10 rupees coins, causing concern among people in India about what to do with them. However, the government did not announce any such ban on 10 rupees coin, it was simply a baseless rumour. After confirmation from the Reserve Bank of India (RBI), people calmed down, and acceptance of the 10 rupees coins resumed.

2.EXISTING SYSTEM

There are various models which exist for Real &Fake news Detection. The most

prevalent system consists of a model that detects fake news based on keywords as well as the headlines, simultaneously.Passive Aggressive detects fake news using keyword analysis and headline,addressing topic-specific tendencies and author behavior and it contains the sentiment analysis.

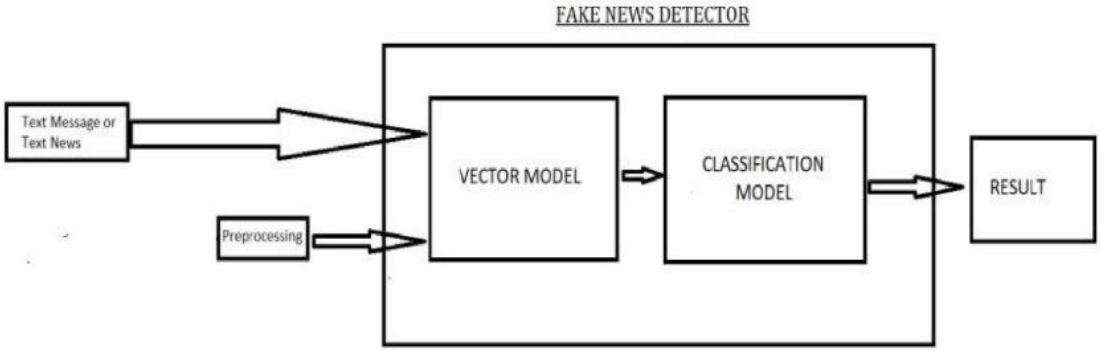


Fig 1 : Fake news detector(Existing)

Disadvantages:

- Limited Contextual Understanding
- It has limitations in Sentiment Analysis
- The concept can be manipulated easy

3. LITERATURE SURVEY:

Yang et al.,[11] TI-CNN model is used for identifying fake news in social media, which performed several methods with accuracy of 92.20%.The dataset collected before the election was held in 2016 US presidential elections.

Patel et al., [12] introduce a Natural Language Processing technique with different classifiers to detect whether the news is real or fake .Algorithms like SVM and KNN gave results with an accuracy of 88.47% and 86.90% while K-means gave results with low accuracy 40.37%.

Kulkarni et al., [5] Their work on the classifiers like Random Forest, Logistic Regression, Decision Tree and Gradient Boosting Algorithms. Logistic Regression accomplish the highest accuracy of 85.04%, followed by Random Forest with 84.50% accuracy and Decision Tree achieve 80.20% ,while Gradient Boosting algorithm accomplish the lowest accuracy of 77.44%.

Agudelo et al.,[10] Detecting False news using machine learning algorithms, natural language processing, and Python programming. By using algorithms like Multinomial Naive Bayes model,CountVectorizer and TF-IDF Vectorizer algorithms, we accomplish the high accuracy of 88.1% and 84.8% on dataset consisting of over 10,000 news items.

4. PROPOSED SYSTEM

In this paper we are going to make use of Natural Language Techniques to overcome the widespread of false news on the internet. Here we make use techniques to determine how the Multinomial algorithm works on the given clip of information which can be given as input to the system.

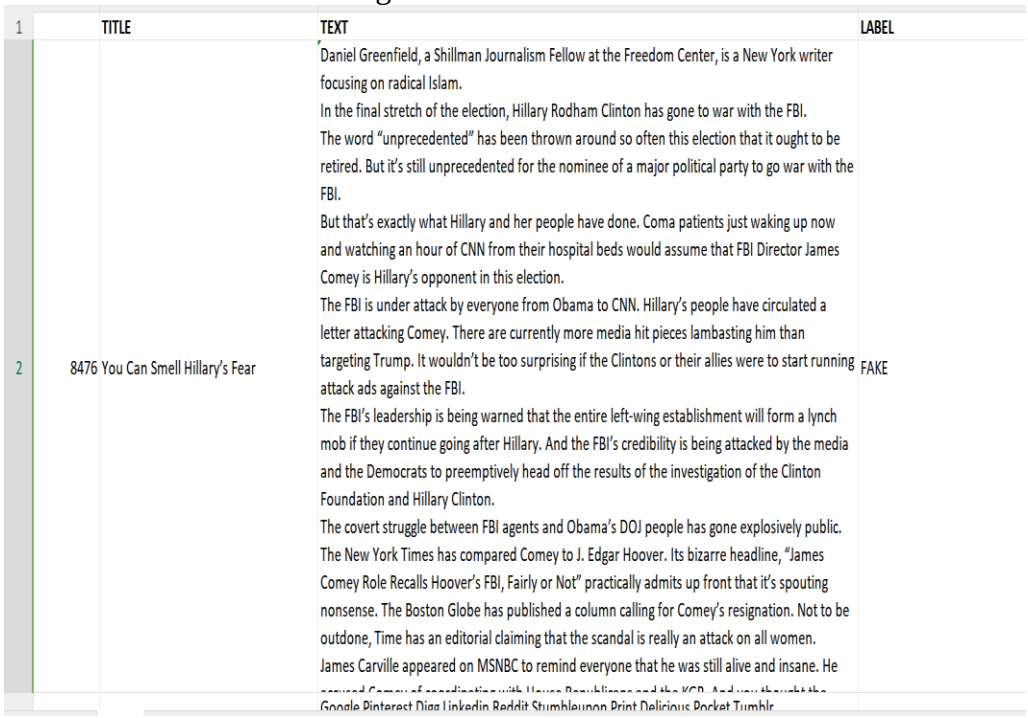
The approach used in this project is to first train the system and then add the news information for which one needs to check if its reliable or not reliable as well as print the accuracy of the algorithm performance on the news clip inserted by the respective reader.

4.1: METHODOLOGY:

We choose the MultinomialNB Classifier because, it performs satisfactory with data sets with high dimensionality and it’s mainly particular classifier when comes to the text

classification.Multinomial Naive Bayes assumes a feature vector where each element represents the number of times it appears (or, very often, its frequency).

The dataset we'll use for this project- we'll call it news.csv. This dataset has a shape of 7796×4. The first column identifies the news, the second and third are the title and text, and the fourth column has labels denoting whether the news is REAL or FAKE.



		TITLE	TEXT	LABEL
2	8476	You Can Smell Hillary's Fear	Daniel Greenfield, a Shillman Journalism Fellow at the Freedom Center, is a New York writer focusing on radical Islam. In the final stretch of the election, Hillary Rodham Clinton has gone to war with the FBI. The word "unprecedented" has been thrown around so often this election that it ought to be retired. But it's still unprecedented for the nominee of a major political party to go war with the FBI. But that's exactly what Hillary and her people have done. Coma patients just waking up now and watching an hour of CNN from their hospital beds would assume that FBI Director James Comey is Hillary's opponent in this election. The FBI is under attack by everyone from Obama to CNN. Hillary's people have circulated a letter attacking Comey. There are currently more media hit pieces lambasting him than targeting Trump. It wouldn't be too surprising if the Clintons or their allies were to start running attack ads against the FBI. The FBI's leadership is being warned that the entire left-wing establishment will form a lynch mob if they continue going after Hillary. And the FBI's credibility is being attacked by the media and the Democrats to preemptively head off the results of the investigation of the Clinton Foundation and Hillary Clinton. The covert struggle between FBI agents and Obama's DOJ people has gone explosively public. The New York Times has compared Comey to J. Edgar Hoover. Its bizarre headline, "James Comey Role Recalls Hoover's FBI, Fairly or Not" practically admits up front that it's spouting nonsense. The Boston Globe has published a column calling for Comey's resignation. Not to be outdone, Time has an editorial claiming that the scandal is really an attack on all women. James Carville appeared on MSNBC to remind everyone that he was still alive and insane. He	FAKE

Fig 2 : Dataset Image

The Proposed system is a application means to classify whether the article is fake or real using Natural Language Processing techniques and machine learning.

We create a user-friendly web interface and there the users can give input url of news article to check whether it is fake or real. When comes to backend system built with using flask, and it is a python web framework. Flask web server handles requests from interface and then it processes them and returns the classification results to the user.

When it comes to NLP modules we use mainly newspaper3k and pre-processing. Firstly, we discuss the concept called Newspaper3k .It is used for web scraping and extracts the news article content from provided URL. Secondly, we discuss the concept called Pre-processing .When the newspaper3k extract the content and gives to the pre-processing and it undergoes the steps like tokenization, stopword removal, and possibly stemming or lemmatization to prepare it for analysis.

After, prep-processing the text data is converted into numerical features that can be understood by the MultinomialNB model. Some techniques like word embedding or TF-IDF may be employed for this purpose. The pre-processed features are boarded into a machine learning model trained on a labeled dataset of articles. There several machine learning models are present for classification but we choose the MultinomialNB classifier and the model predicts whether the article is fake or real based on the extracted features.

Overall, the system contains NLP techniques and machine learning models to automatically classify news articles, thereby users identify the misleading or false information.

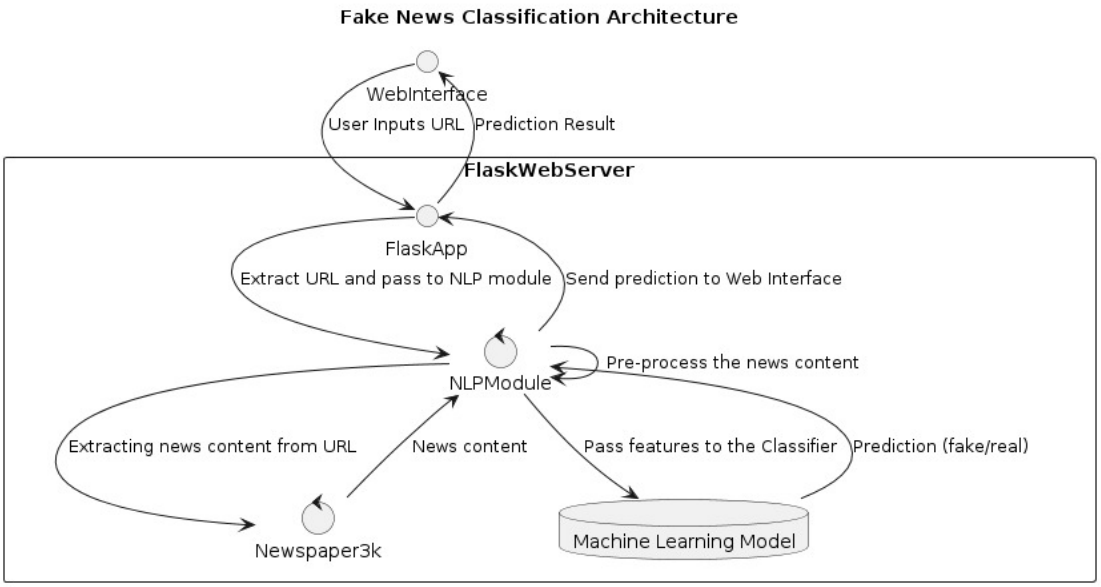


Fig 3: Fake news classification using nlp architecture

4.2: MULTINOMIAL NAIVE BAYES ALGORITHM:

Multinomial Naive Bayes is a machine learning algorithm that is based on bayes' theroem.It is a probilistic classifier to calculate the probability distribution of given data which in the form of text,which makes it suited for data and the features are represent discrete frequencies or count of events in various natural language processing(NLP) tasks.The probability mass function (PMF) of the Multinomial distribution is used to model the likelihood of observing a specific set of word counts in a document. It is given by:

$$P(D|c) = \frac{T_c!}{\prod_{i=1}^{V}(x_i!)} \prod_{i=1}^{V} \left(\frac{\theta_{c,i}^{x_i}}{x_i!} \right)$$

4.3: IMPLEMENTATION

- Loading Data:The cleaned dataset news.csv containing text and label columnsand the shape of the data set is 7796×4 and it contains attributes like title,text and label.
- Splitting Data:The dataset is splits in the ratio of 8:2 that means 80% of the data is training purpose and 20% of the data is testing purpose.
- Creating Pipelene:We create a machine learning pipeleine that applis TF-IDF vectorization to the text data to convert it into numerical features and then applies the Multinomial Naive Bayes Classifier.
- Training the Model:Train the pipeline on training data(X_train and y_train).
- Predicting Labels:Using trained model we predict the labels for test data.
-Model Evaluation:Evaluate the performance using confusion matrix and classification report.
- Deployment:Deployment were held in the user interface using environment which allows users to identify where the news is real or fake.

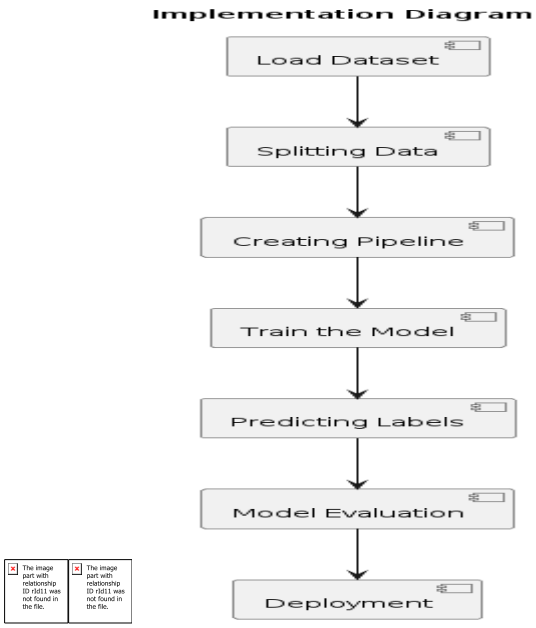


Fig 4:- Implementation

5. RESULTS AND DISCUSSION:

In this project we predict whether the news using the NLP module and we use the fake and true news datasets for creation of this system.We used to performance Text Pre-processing and Vectorization to detect the news.we obtained the accuracy of 85.0%,hence we declared the news is real or fake.

6.SAMPLE SCREENSHOTS:

In this below screenshot we see the front end page and we just fed the news url link then it gives the news whether the news is fake or real.

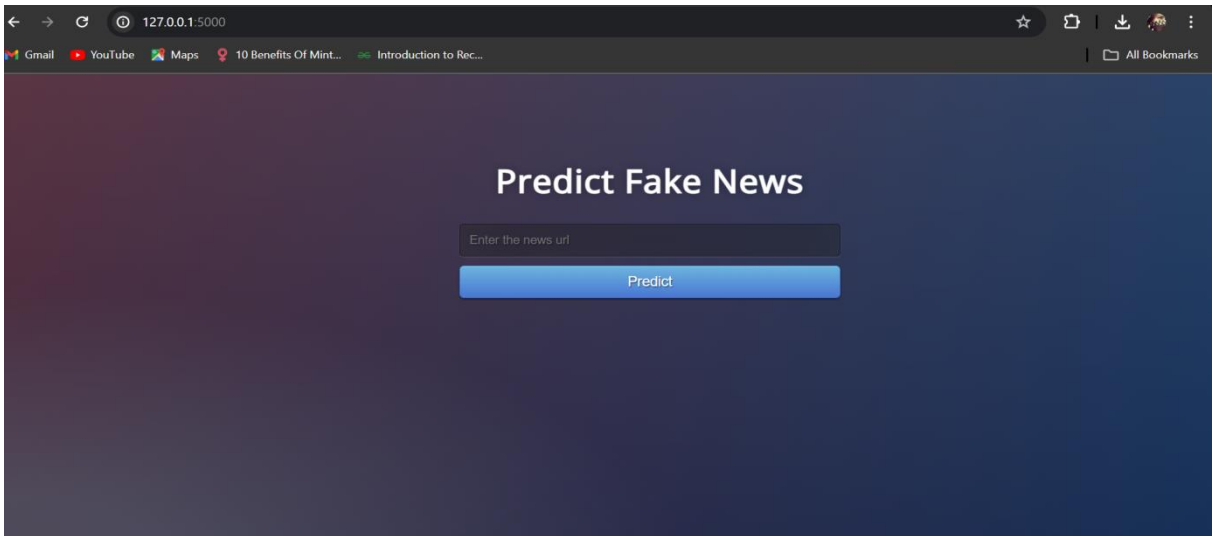


Fig 5:- Front-end page without article link

In the below Screenshot we see that the link provided and it goes to NLP module and then the newspaper3k is extracts the content and we apply the content to the machine learning model and it goes to module by using flask.

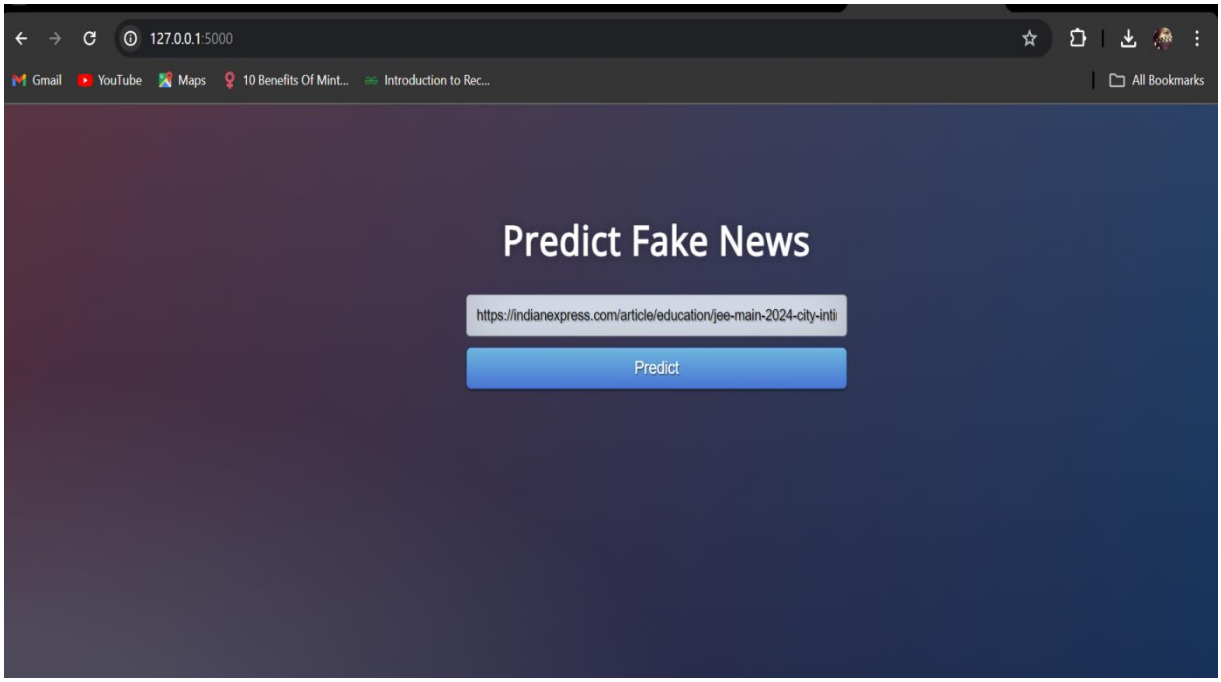


Fig 6:- Front-end page with article link

In the below Screenshot we see that the news is classify by using the newspaper3k andit extract the content and apply preprocessing,machine learning models and b using flaskwebserver we get the prediction and it tells whether the article news is real or fake.

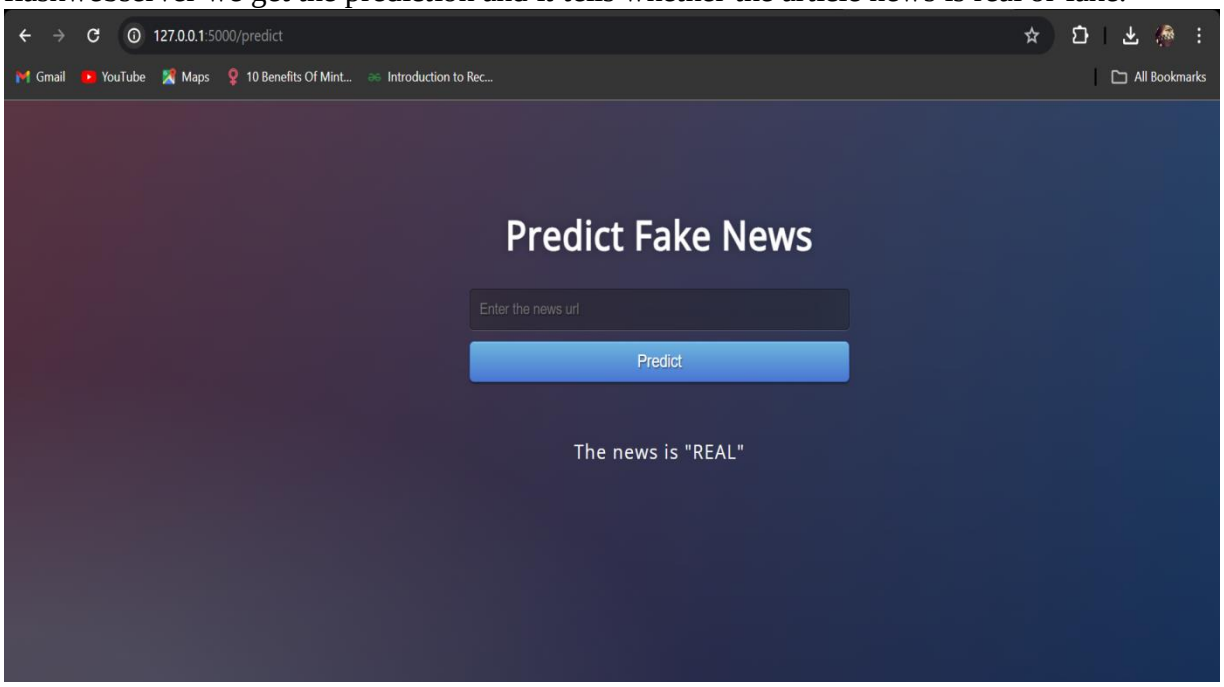


Fig 7:- we get whether news is real or fake

7. CONCLUSION:

Fake news responsible for creating false and misleading information that greatly affect the people and the event. This project explains what fake news are and what real news are by using Natural Language Processing and Machine Learning model which is used for classification. We use NLP for automatically predict and detect the news whether it is real or fake news. In This project develop a web application for fake news classification using

Natural Language Processing techniques. We use flask for back-end purpose and we allows the users to give input news article URLs. The application get content by using newspaper3k, process it using a pre-trained ML model, and returns a classification result. The front -end displays the result.
Overall, the project aims to oppose misinformation by providing the tool to identify whether the news is real or fake potentially.

8. FUTURE SCOPE:

In future of fake news classification using nlp ,several chances for exploration and enhancement. This could involve experimenting with different machine learning algorithms or we can say that nlp techniques like feature engineering to improve model accuracy and efficiency. There is chance to explore different sources, languages and types of news to make your model more robust and adaptable to various contexts. Focusing on these points, they can continue to evolve and making it more effective, reliable and valuable to users.

9.REFERENCES:

[1] Chowdhury, G. (2003) Natural language processing. Annual Review of Information Science and Technology, 37. pp. 51-89. ISSN 0066-4200
[2] A. N. K. Movanita, "BIN: 60 Persen Konten Media Sosial adalah Informasi Hoaks (BIN: 60 percent of social media content ishoax)," 2018. [Online]. Available: https://nasional.kompas.com/read/2018/03/15/0647555 1/bin-60- persen-konten-media- socialadalah-informasi- hoaks.
[3] S. Kumar, R. Asthana, S. Upadhyay, N. Upreti, and M. Akbar, "Fake news detection using deep learning models: A novel approach," Transactions on Emerging Telecommunications Technologies, 2019.
[4] K.-H. Choi, "A study on the effect of reading side tool with NLP skill on student chinese reading performance," Master Thesis, Grad. Ins. Edu. Inf. and Meas., National Taichung Univ. of Edu., Taichung, Taiwan, 2015.
[5] P. Kulkarni, S. Karwande, R. Keskar, P. Kale, and S. Iyer, “Fake News Detection using Machine Learning,” ITM Web of Conferences, vol. 40, p. 03003, 2021, doi: 10.1051/itmconf/20214003003.
[6] Raj Bridgelall Department of Transportation, Logistics, and Finance, College of Business, North Dakota State University, Fargo, ND 58108, USA; raj.bridgelall@ndsu.edu
[7] Marwan Omar, Soohyeon Choi, DaeHun Nyang, and David Mohaisen, 3 Jan 2022
[8] Fake news detector: NLP project by ishant juyal
(https://levelup.gitconnected.com/fake-news-detector-nlp-project-9d67e0177075)
[9]Aldwairi, M. and A. Alwahedi, Detecting Fake News in Social Media Networks.
[10] G. Agudelo, O. Parra, and J. Barón Velandia, “Raising a Model for Fake News Detection Using Machine Learning in Python,” pp. 596–604, 2018, doi: 10.1007/978-3-030-02131-3_52ï.
[11] . Y. Yang et al., “TI-CNN: Convolutional Neural Networks for Fake News Detection,” Jun. 2018, Accessed: Mar. 24, 2023. [Online]. Available: https://arxiv.org/abs/1806.00749v3.
[12] J. Patel, M. Barreto, U. Sahakari, and Dr. S. Patil, “Fake News Detection with Machine Learning,” International Journal of Innovative Technology and Exploring Engineering, vol. 10, no. 1, pp. 124–127, Nov. 2020, doi: 10.35940/IJITEE.A8090.1110120
[13]H. Allcott, M. Gentzkow, and C. Yu, “Trends in the diffusion of misinformation on social media,” Research and Politics, vol. 6, no. 2, Apr. 2019, doi: 10.1177/2053168019848554.
[14] S. M. Jones-Jang, T. Mortensen, and J. Liu, “Does Media Literacy Help Identifi cation of Fake News? Information Literacy Helps, but Other Literacies Don’t,”American Behavioral Scientist, vol. 65, no. 2, pp. 371–388, Feb. 2021, doi: 10.1177/0002764219869406.
[15] P. Machete and M. Turpin, “The Use of Critical Thinking to Identify Fake News: A Systematic Literature Review,” Lecture Notes in Computer Science (including subseries Lecture Notes in Artifi cial Intelligence and Lecture Notes in Bioinformatics), vol. 12067 LNCS, pp. 235–246, 2020, doi: 10.1007/978-3-030-45002-1_20.