KINGS ENGINEERING COLLEGE
Chennai-Bangalore Highway, Irungattukottai, Sriperumbudur, Chennai - 602 117.
Ph: 044 - 71224401 – 08. Fax: 044 - 71224410
ACCREDITED WITH NAAC, APPROVED BY AICTE & AFFILIATED TO ANNA UNIVERSITY

Department of Artificial Intelligence & Data Science

College Code: 2108

Project Domain: Artificial Intelligence

Project Title : Fake News Detection Using NLP

Project Mentor: Mr.V.Sundarraj

Team Members:

1) Vignesh C(210821243062)
2) Kishore R(210821243034)
3) Manoj Kumar L(210821243075)
4) Om Aravind L(210821243034)

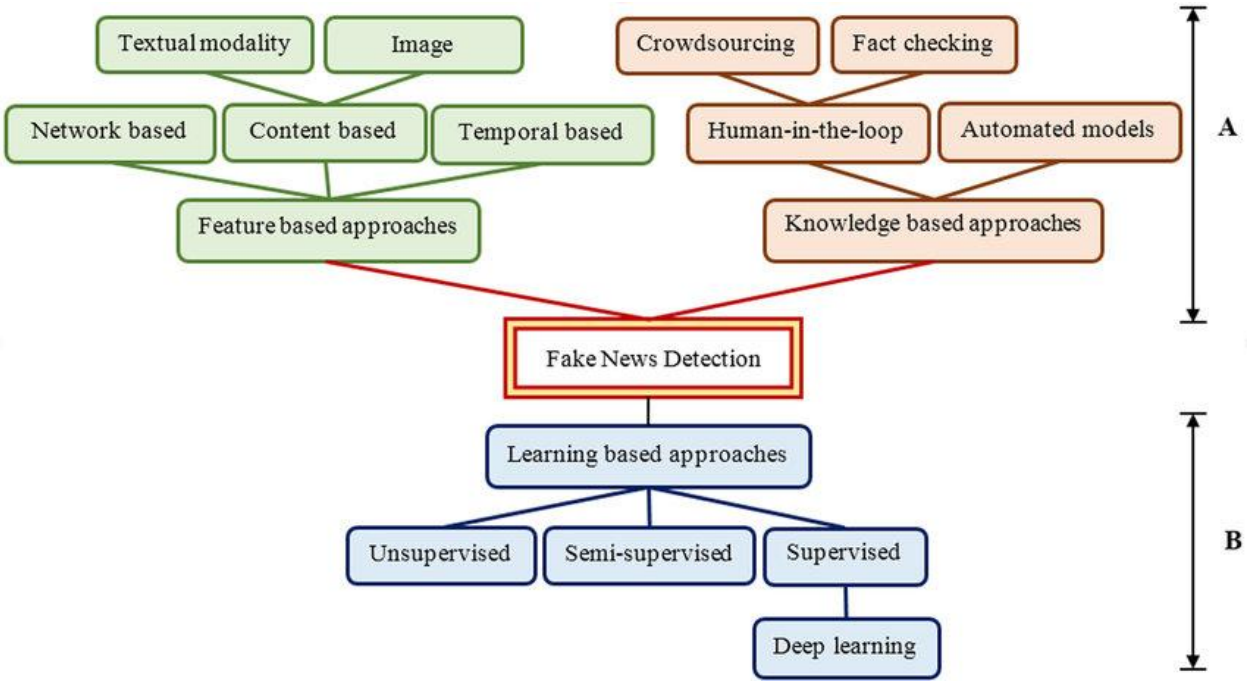
Abstract:

This study explores the application of Natural Language Processing (NLP) techniques to enhance the detection of fake news. Leveraging linguistic patterns, sentiment analysis, and source credibility assessment, our approach aims to empower automated systems to discern misinformation from legitimate news sources. The proposed methodology showcases promising results in improving the accuracy and efficiency of fake news identification, contributing to the ongoing efforts in combating the spread of deceptive information in the digital age.

Problem Statement:

“Fake News Detection Using NLP”

Design Thinking Process:



Phases of Development:

1. Conceptualization.
2. Planning.
3. Implementation.
4. Testing.
5. Deployment.
6. Maintenance and Iteration.

Dataset Used:

	A	B	C	D	E
1	title	text	subject	date	
2	As U.S. budg	WASHINGT(	politicsNew	31-Dec-17	
3	U.S. military	WASHINGT(	politicsNew	29-Dec-17	
4	Senior U.S.	WASHINGT(	politicsNew	31-Dec-17	
5	FBI Russia p	WASHINGT(	politicsNew	30-Dec-17	
6	Trump want	SEATTLE/W	politicsNew	29-Dec-17	
7	White Hous	WEST PALM	politicsNew	29-Dec-17	
8	Trump says	WEST PALM	politicsNew	29-Dec-17	
9	Factbox: Tr	The followir	politicsNew	29-Dec-17	
10	Trump on T	The followir	politicsNew	29-Dec-17	
11	Alabama of	WASHINGT(	politicsNew	28-Dec-17	
12	Jones certif	(Reuters) - A	politicsNew	28-Dec-17	
13	New York g	NEW YORK/	politicsNew	28-Dec-17	
14	Factbox: Tr	The followir	politicsNew	28-Dec-17	
15	Trump on T	The followir	politicsNew	28-Dec-17	
16	Man says he	(In Dec. 25	politicsNew	25-Dec-17	
17	Virginia offi	(Reuters) - A	politicsNew	27-Dec-17	
18	U.S. lawmal	WASHINGT(	politicsNew	27-Dec-17	
19	Trump on T	The followir	politicsNew	26-Dec-17	
20	U.S. appeal	(Reuters) - A	politicsNew	26-Dec-17	
21	Treasury Se	(Reuters) - A	politicsNew	24-Dec-17	
22	Federal judg	WASHINGT(	politicsNew	24-Dec-17	
23	Exclusive: U	NEW YORK	politicsNew	23-Dec-17	
24	Trump trave	(Reuters) - A	politicsNew	23-Dec-17	
25	Second cou	WASHINGT(	politicsNew	23-Dec-17	
26	Failed vote	LIMA (Reute	politicsNew	23-Dec-17	
27	Trump signs	WASHINGT(	politicsNew	22-Dec-17	
28	Companies	WASHINGT(	politicsNew	23-Dec-17	
29	Trump on T	The followir	politicsNew	22-Dec-17	
30	Mexico to r	MEXICO CIT	politicsNew	22-Dec-17	

Data preprocessing:

1. Handling Missing Data.
2. Removing Duplicates .
3. Data Transformation.
4. Outlier Detection and Handling.
5. Normalization/Standardization.

Feature Extraction Techniques:

1. Principal Component Analysis (PCA): A dimensionality reduction technique that transforms the data into a new coordinate system, capturing the most significant variations.
2. Feature Scaling: Ensure that all features are on a similar scale. This is often part of preprocessing but is worth mentioning here.
3. One-Hot Encoding: Convert categorical variables into binary vectors. Each category becomes a new binary feature.
4. Text Vectorization: If dealing with textual data, techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or word embeddings (like Word2Vec or GloVe) can be used to convert text into numerical vectors.
5. Feature Engineering: Creating new features based on existing ones. For example, combining "height" and "weight" to create a "body mass index" feature.

Choice of Classification Algorithm:

1. Logistic Regression.
2. Decision Trees.
3. Support Vector Machines (SVM).
4. Naive Bayes.
5. K-Nearest Neighbors (KNN).
6. Neural Networks.

Model Training Process:

1. Feature Scaling: Ensure that your features are on a similar scale. This is particularly important for algorithms that rely on distances between data points, like KNN or SVM.
2. Model Initialization: Set up the chosen algorithm with initial parameters.
3. Model Training: Present your training data to the model. The algorithm adjusts its parameters based on the input data and their corresponding labels.
4. Model Evaluation: Use the testing set to evaluate how well your model generalizes to new, unseen data. Common metrics include accuracy, precision, recall, and F1 score.
5. Hyperparameter Tuning: Adjust the hyperparameters of your model to improve its performance. This may involve techniques like grid search or random search.
6. Model Deployment: Once satisfied with the model's performance, deploy it for making predictions on new, real-world data.