IBM NAAN MUDHALVAN

ARTIFICIAL INTELLIGENCE-GROUP 3

PROJECT:

FAKE NEWS DETECTON USING NLP

TEAM MEMBER :

NAME: SREELAKSHMI.M

REG NO:721921243109

PHASE 2:INNOVATION



In this phase,exploring advanced techniques like deep learning can significantly improve fake news detection accuracy. Deep learning models, such as convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have shown promise in this domain. In this phase submission I explain some ways in which they can enhance accuracy:

**Feature Extraction:**

Deep learning models can automatically learn relevant features from text and metadata, reducing the need for manual feature engineering. This can capture subtle patterns in fake news that might be missed by traditional methods.

**Text Embeddings:**

Pre-trained word embeddings like Word2Vec, GloVe, or BERT can be used to represent words in a more semantically meaningful way, aiding in understanding the context and nuances of news articles.

**Deep Neural Networks:**

Models like LSTM and GRU are effective for sequential data, making them suitable for analyzing the temporal aspect of news articles and social media posts, which can help in identifying deceptive narratives.

**Ensemble Learning:**

Combining multiple deep learning models through ensemble techniques like stacking or bagging can improve overall detection accuracy. Each model may capture different aspects of fake news.

**Multimodal Approaches:**

Integrating information from various sources, such as text, images, and user profiles, using deep learning architectures can enhance accuracy, as fake news often involves diverse content types.

**Transfer Learning :** Fine-tuning pre-trained deep learning models on fake news detection tasks can leverage knowledge learned from large text corpora, further boosting performance.

**Continuous Learning:**

Implementing online learning with deep models allows systems to adapt to evolving fake news tactics, staying effective over time.

**Data Augmentation:**

Using data augmentation techniques can help address the scarcity of labeled data by generating synthetic examples for training, reducing overfitting.

**Adversarial Training:**

Training models to recognize adversarial attacks and generate robust features can make them more resilient to sophisticated fake news generators.

Dataset Link:

https://www.kaggle.com/datasets/clmentbisaillon/fake-and-real-news- dataset

SOURCE CODE:

import pandas as pd

from sklearn.model\_selection import train\_test\_split

from sklearn.feature\_extraction.text import TfidfVectorizer

from sklearn.linear\_model import LogisticRegression

from sklearn.metrics import accuracy\_score, confusion\_matrix, precision\_score, recall\_score

fake\_data = pd.read\_csv('Fake.csv')

fake\_data['label'] = 'FAKE'

true\_data = pd.read\_csv('True.csv')

true\_data['label'] = 'REAL'

data = pd.concat([fake\_data, true\_data], ignore\_index=True)

X = data['title'] + ' ' + data['text'] + ' ' + data['subject'] + ' ' + data['date']

y = data['label']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

tfidf\_vectorizer = TfidfVectorizer(stop\_words='english')

X\_train\_tfidf = tfidf\_vectorizer.fit\_transform(X\_train)

X\_test\_tfidf = tfidf\_vectorizer.transform(X\_test)

epochs = 10

model = LogisticRegression(max\_iter=epochs)

model.fit(X\_train\_tfidf, y\_train

y\_pred = model.predict(X\_test\_tfidf)

However, it's essential to consider the ethical implications, biases, and potential challenges when implementing advanced deep learning techniques in fake news detection, as they can be computationally intensive and may require substantial labeled data for effective training. Continuous evaluation and refinement of these models are crucial to stay ahead of evolving misinformation tactics