**NAAN MUDHALVAN PROJECT(IBM)**

IBM AI 101 **ARTIFICIAL INTELLIGENCE-GROUP 1**

PROJECT:

TEAM-6 **FAKE NEWS DETECTION USING NLP**

TEAM MEMBERS

1)GOKUL NATH S (reg no. 110321106013)

2)DINESH K (reg no. 110321106011)

3)NAREANDHIRAN K B (reg no. 110321106032)

4)JEROME BENNY HINNDAS (reg no. 110321106015)

**Problem Statement:**

Design and develop an NLP-based system that can accurately identify and classify news articles or information as either "fake" or "real" by analyzing the textual content, with the primary goal of mitigating the spread of misinformation and promoting the dissemination of trustworthy information.

**Key Components and Objectives:**

**1.** **Data Collection:** Gather a diverse and comprehensive dataset of news articles, comprising both legitimate and fake news, to train and test the model. This dataset should encompass various domains and sources.

**2.** **Text Preprocessing:** Clean and preprocess the textual data, which includes tasks such as tokenization, stemming, stop-word removal, and handling special characters.

**3.** **Feature Extraction:** Utilize NLP techniques to extract meaningful features from the text, such as TF-IDF (Term Frequency-Inverse Document Frequency) vectors or word embeddings (e.g., Word2Vec, GloVe).

**4.** **Model Selection:** Choose an appropriate machine learning or deep learning model for fake news detection. Common choices include supervised learning algorithms like logistic regression, decision trees, random forests, and neural networks (e.g., LSTM, BERT).

**5.** **Training and Validation:** Train the selected model using labeled data and validate its performance using various evaluation metrics, such as accuracy, precision, recall, F1-score, and AUC-ROC.

**6.** **Cross-Validation:** Perform cross-validation to ensure the model's robustness and generalizability by splitting the dataset into training, validation, and test sets.

**7. Feature Engineering:** Experiment with different feature engineering techniques, including n-grams, sentiment analysis, and topic modeling, to improve the model's performance.

**8. Real-time Monitoring:** Implement a system for real-time monitoring of news sources or social media platforms to detect and classify potentially fake news as soon as they emerge**.**

**9. Explainability and Interpretability:** Ensure that the model's predictions can be explained and interpreted, as transparency is essential in building trust.

**10. Deployment:** Deploy the trained model in a user-friendly interface or integrate it into news platforms and social media networks to flag or remove potentially fake news.

**11. Continuous Learning:** Implement mechanisms for continuous model retraining and adaptation to evolving fake news tactics and trends**.**

**12. Evaluation and Feedback Loop:** Continuously evaluate the model's performance and collect user feedback to make necessary improvements.

The ultimate goal of this NLP-based fake news detection system is to contribute to the reduction of the spread of misinformation and the promotion of accurate and reliable information in the digital space.