**FAKE NEWS DETECTION USING NLP**

TEAM MEMBER

810021106030: HARINI V

PHASE 1: DOCUMENT SUBMISSION

**PROJECT DEFINITION:**

The goal of this project is to develop a system that uses Natural Language Processing (NLP) techniques to automatically identify and classify fake news articles from legitimate news sources. Fake news has become a significant problem in the age of digital media, and this project aims to contribute to the efforts in combating the spread of misinformation.To get the accurately classified collection of news as real or fake we have to build a machine learning model.

**DESIGN THINKING:**

**Data Collection and Preparation:**

* Acquire a diverse dataset of news articles, including both genuine and fake news sources.
* Clean and preprocess the use text data, including tasks such as tokenization, stop word removal, and stemming/lemmatization.

**Feature Extraction:**

* Utilize NLP techniques to extract relevant features from the text data.
* Features may include word embeddings (e.g., Word2Vec or GloVe), TF-IDF vectors, and linguistic features (e.g., sentiment analysis, readability scores).

**Data Splitting:**

* Divide the dataset into training, validation, and testing sets to train and evaluate the machine learning models.

**Model Development:**

* Implement and experiment with various machine learning and deep learning models for fake news detection. Common choices include:
  + - * Logistic Regression
      * Naive Bayes
      * Random Forest
      * LSTM or Transformer-based models (e.g., BERT)
* Fine-tune models on the training data and validate their performance using the validation set.

**Model Evaluation:**

* Assess the models’ performance using appropriate evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC.
* Employ cross-validation to ensure the model’s robustness.

**Testing and Validation:**

* Conduct thorough testing to ensure the system’s reliability and accuracy in real-world scenarios.

**Continuous Improvement:**

* Monitor the system’s performance over time and update it with new data and improvements to stay effective in detecting evolving fake news tactics.

By implementing these modules, this project aims to provide a robust and reliable solution for fake news detection, helping users make more informed decisions when consuming news content.