IBM NAN MUDHALVAN PROJECT

**FAKE NEWS DETECTION USING NLP**

**Problem Definition:**

**Empathise:** Understand the impact of fake news on society, users' need for reliable information, and the challenges of identifying fake news from genuine sources**.**

**Define:** Formulate a clear problem statement: "Develop an NLP-based system to detect fake news in news articles and social media posts, providing users with accurate and reliable information".

**Design Thinking for Fake News Detection:**

**Data Source:**

**Empathise:** Understand the types of data needed for fake news detection (textual content, metadata, user engagement).

**Ideate:** Explore potential data sources, such as news websites, social media platforms, and fact-checking organisations.

**Prototype**: Create data scrapers or APIs to collect and preprocess data from these sources**.**

**Data Processing:**

**Empathise:** Recognize the challenges of handling noisy and unstructured textual data.

**Ideate**: Design a data preprocessing pipeline that includes text cleaning, tokenization, stop-word removal, and stemming/lemmatization.

**Prototype**: Develop scripts or workflows for data preprocessing.

**Model Selection:**

**Empathise:** Understand the importance of selecting appropriate NLP models for fake news detection.

**Ideate:** Consider various NLP models such as TF-IDF, Word Embeddings (e.g., Word2Vec, GloVe), and pre-trained language models (e.g., BERT, GPT).

**Prototype:** Experiment with different models to understand their performance.

**Model Training:**

**Empathise:** Recognize the need for labelled training data (real and fake news samples).

Ideate: Create a labelled dataset by combining manually labelled examples and existing datasets like LIAR-PLUS, FakeNewsNet, or Snopes.

**Prototype:** Train NLP models on the labelled dataset using appropriate algorithms and techniques.

**Evaluation:**

**Empathise:** Understand the importance of evaluation metrics in assessing model performance.

**Ideate:** Choose evaluation metrics like precision, recall, F1 score, and accuracy.

**Prototype:** Evaluate the models using a holdout test dataset and fine-tune them to optimise the chosen metrics.

**Deployment:**

**Empathise:** Recognize the need to integrate the solution into users' daily online experience.

**Ideate**: Develop user-friendly browser extensions or mobile apps for real-time fake news detection, or integrate the solution into social media platforms.

**Prototype:** Create a user interface that provides immediate feedback on news articles and social media posts.

**Monitoring and Maintenance:**

**Empathise:** Understand the importance of continuous model updates and monitoring.

**Ideate**: Set up monitoring systems to track model performance and adapt to emerging fake news trends.

**Prototype:** Implement regular model retraining and updates to ensure effectiveness overtime.

**Education and Awareness:**

**Empathise:** Recognize the role of user education in combating fake news.

**Ideate:** Develop educational content and resources to teach users about fake news, critical thinking, and fact-checking.

**Prototype:** Create a section within your solution for users to access educational materials.