AI – PHASE 2
ARTIFICIAL INTELLIGENCE
[FAKE NEWS DETECTION USING NLP]

IBM

TEAM MEMBERS

1. **BHARANI DHARAN .M[LEADER] -- 731221104302**
2. **SURYA . S 731221104036**
3. **ARULMURUGAN .S 731221104004**
4. **VASANTH KUMAR . S 731221104038**
5. **SRINIVAS . V 731221104307**

AGENDA

- ABSTRACT
- FAKE NEWS DETECTION
- NATURAL LANGUAGE PROCESSING
- DATA COLLECTION
- PREPROCESSING
- FEATURE EXTRACTION
- MODEL SELECTION
- TRAINING
- EVALUATION

ABSTRACT

➢In the age of digital media, fake news is a serious and harm problem because it spreads misinformations individuals, organizations, and even entire nations which is a challenging aspect.

➢ This study proposes a using python(NLP) approach for detecting fake news.

FAKE NEWS DETECTION

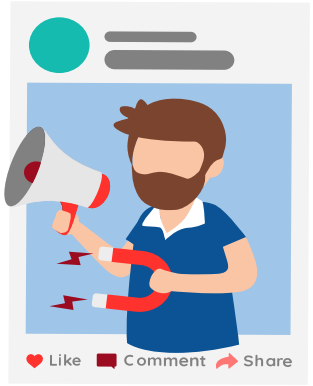
- Fake news is a real problem in today's world, and it has become more extensive and harder to identify
- A major challenge in fake news detection is to detect it in the early phase.
- Another challenge in fake news detection is the unavailability or the shortage of labeled data for training the detection models.
- We propose a novel fake news detection framework that can address these challenges.
- Our proposed framework exploits the information from the news articles and the social contexts to detect fake news.

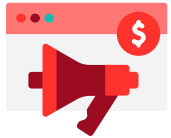
Natural Language Processing (NLP)

➢Utilize NLP techniques to analyze the language, grammar, and style of news articles.

➢Look for patterns that are indicative of fake news, such as sensationalism, extreme language, or excessive use of exclamation points.

➢Train machine learning models on labeled datasets to classify news articles as real or fake based on linguistic features.





DATA COLLECTION

- Gather a diverse dataset of news articles that includes both real and fake news.
- Ensure that the dataset is well-labeled to serve as the ground truth.

PREPROCESSING

- Tokenization: Split the text into words or tokens.
- Stop word removal: Eliminate common words like "and," "the," and "is" that may not contribute much to the classification task.
- Text cleaning: Remove any HTML tags, special characters, punctuation, and other irrelevant elements from the text.
- Lemmatization or stemming: Reduce words to their base or root form.

FEATURE EXTRACTION

- Convert the text data into numerical features that can be used for machine learning.
- Common methods include TF-IDF (Term Frequency-Inverse Document Frequency)
- Word embeddings (Word2Vec, GloVe).

MODEL SELECTION

- **Model Selection**:
- Choose an appropriate machine learning model for classification. Common choices include:
  - **Logistic Regression**
  - **Naive Bayes**
  - **Random Forest**
  - **Support Vector Machine (SVM)**
  - **Deep Learning models (e.g., LSTM, BERT)**

TRAINING

- Split your dataset into training and validation sets.
- Train your chosen model on the training data. Fine-tune hyper parameters for optimal performance.

EVALUATION

- Use evaluation metrics such as accuracy, precision, recall, F1-score, and ROC-AUC to assess the model's performance.
- Be cautious of class imbalance, as fake news is often a minority class. You may need to consider techniques like oversampling, under sampling, or using different evaluation metrics.

THANK YOU!