

# FAKE NEWS DETECTION USING NPL

## 1. Data Collection:

- Gather a dataset of labeled news articles, where each article is classified as either "real" or "fake."

## 2. Data Preprocessing:

- Text cleaning: Remove punctuation, special characters, and unnecessary whitespace.
- Tokenization: Split the text into individual words or tokens.
- Text vectorization: Convert text data into numerical format using techniques like TF-IDF or word embeddings (Word2Vec, GloVe, etc.).

## 3. Feature Extraction:

- Generate relevant features from the text data. Features can include word frequency, n-grams, and sentiment analysis.

## 4. Model Development:

- Choose a machine learning or deep learning model for fake news detection. Common models include:
  - Multinomial Naive Bayes
  - Logistic Regression
  - Random Forest
  - Recurrent Neural Networks (RNNs)
  - Convolutional Neural Networks (CNNs)
  - Bidirectional LSTMs
- Split the dataset into training and testing sets.

## **5. Model Training:**

- Train your selected model on the training data.

## **6. Model Evaluation:**

- Evaluate your model using metrics like accuracy, precision, recall, F1-score, and confusion matrix on the testing dataset.
- Use k-fold cross-validation for a more robust evaluation.

## **7. Hyperparameter Tuning:**

- Optimize your model by adjusting hyperparameters, such as learning rates, batch sizes, and network architecture.

## **8. Model Deployment (Optional):**

- If you plan to deploy the model in a real-world application, save the trained model and create an API or a web interface.

## **9. Project Documentation:**

- Create documentation that explains the project, the dataset, the model, and the results.

## **10.Submission Preparation:**

- Organize your project code and documentation for submission, including a README file with clear instructions.

## **11.Future Work and Conclusion:**

- Discuss possible improvements, limitations, and future work for your fake news detection project.

## **PROGRAM;**

The Accuracy of your fake news detection system.

```
import pandas as pd
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
from sklearn.naive_bayes import MultinomialNB
```

```
from sklearn.metrics import accuracy_score, classification_report
```

```
data = pd.read_csv("fake_news_dataset.csv")
```

```
data['text'] = data['text'].str.lower()
```

```
data['text'] = data['text'].str.replace('[^\w\s]', '')
```

```
data['text'] = data['text'].str.strip()
```

```
X = data['text']
```

```
y = data['label']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,  
random_state=42)
```

```
vectorizer = TfidfVectorizer(max_features=5000)
```

```
X_train_tfidf = vectorizer.fit_transform(X_train)
```

```
X_test_tfidf = vectorizer.transform(X_test)
```

```
classifier = MultinomialNB()
```

```
classifier.fit(X_train_tfidf, y_train)
```

```
y_pred = classifier.predict(X_test_tfidf)
```

```
accuracy = accuracy_score(y_test, y_pred)
```

```
report = classification_report(y_test, y_pred)
```

```
print(f"Accuracy: {accuracy}")
```

```
print(report)
```

OUTPUT ;

CHECK THE ACCURACY IN FAKE NEWS DETECTION

DATASET PROCESSING ;

```
import pandas as pd
```

```
from sklearn.feature_extraction.text import TfidfVectorizer
```

```
from sklearn.model_selection import train_test_split
```

```
from sklearn.naive_bayes import MultinomialNB

from sklearn.metrics import accuracy_score, classification_report

dataset = pd.read_csv("your_dataset.csv")

dataset['text'] = dataset['text'].str.lower() # Convert text to lowercase

dataset['text'] = dataset['text'].str.replace('[^\w\s]', '')

dataset['text'] = dataset['text'].str.strip()

X = dataset['text']

y = dataset['label'] # 'label' is the column containing binary labels (0 for
real, 1 for fake)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

vectorizer = TfidfVectorizer(max_features=5000)

X_train_tfidf = vectorizer.fit_transform(X_train)

X_test_tfidf = vectorizer.transform(X_test)

classifier = MultinomialNB()

classifier.fit(X_train_tfidf, y_train)
```

```
y_pred = classifier.predict(X_test_tfidf)
```

```
accuracy = accuracy_score(y_test, y_pred)
```

```
report = classification_report(y_test, y_pred)
```

GET OUTPUT ;

```
print("Fake News Detection Results:")
```

```
print(f"Accuracy: {accuracy}")
```

```
print(report)
```

**Dataset Link:** <https://www.kaggle.com/datasets/clmentbisailon/fake-and-real-news-dataset>

S.GOPINATH

NANDHA COLLEGE OF TECHNOLOGY

30.10.2023

PHASE 5