

AI-based fake news classifier



PRESENTED BY

STUDENT NAME: CHITRANSH MATHUR

COLLEGE NAME: JAIPUR NATIONAL UNIVRSITY

DEPARTMENT: SCHOOL OF LIFE AND BASIC SCIENCE

EMAIL ID: CHITRANSHMATHUR16@GMAIL.COM

AICTE STUDENT ID: STU67e425e90603d1743005161

OUTLINE

- **Problem Statement** (Should not include solution)
- **Proposed System/Solution**
- **System Development Approach** (Technology Used)
- **Algorithm & Deployment**
- **Result (Output Image)**
- **Conclusion**
- **Future Scope**
- **References**

PROBLEM STATEMENT

In the digital age, misinformation spreads faster than facts. Social media platforms and news websites are often plagued with fake news, which can influence public opinion, cause panic, or manipulate political outcomes. The challenge is to build a system that can automatically detect fake news in real time using machine learning.

PROPOSED SOLUTION

- This project aims to build an **AI-based fake news classifier** using natural language processing (NLP) and machine learning. The solution will:
- Collect and label real vs fake news datasets
- Preprocess the news data (clean text, remove stop words, etc.)
- Extract important features using NLP techniques like TF-IDF or word embeddings
- Train and evaluate classification models such as Logistic Regression, Naive Bayes, and Random Forest
- Develop a basic web app where users can input a news headline or article to verify its authenticity

SYSTEM APPROACH

- **System Development Approach**
- **System Requirements:**
- Python 3.x
- Jupyter Notebook / VS Code
- Internet connection (for dataset/API access)
- **Libraries Used:**
- pandas, numpy (data handling)
- sklearn (ML algorithms)
- nltk, re (text processing)
- Flask (web deployment)

ALGORITHM & DEPLOYMENT

- **Algorithm Selection:**
We used **Logistic Regression** and **Passive Aggressive Classifier** due to their performance on text classification problems. They are fast, scalable, and work well with sparse data like TF-IDF outputs.
- **Data Input:**
 - News title and body
 - Label: Real or Fake
 - Extracted features: word frequencies, n-grams, sentiment markers
- **Training Process:**
 - Used a labeled dataset from Kaggle ("Fake and Real News Dataset")
 - Performed data splitting (80% train, 20% test)
 - Applied TF-IDF vectorization
 - Trained models using cross-validation and grid search for hyperparameters
- **Prediction Process:**
 - The user inputs text
 - Text is preprocessed and vectorized
 - Model outputs either "Real" or "Fake"

RESULT

- **Accuracy Achieved:**

- ✓ ~95% accuracy on test data using TF-IDF + Passive Aggressive Classifier.

- **Model Performance Metrics:**

- **Precision:** ~0.94

- **Recall:** ~0.96

- **F1-Score:** ~0.95

- (These numbers can slightly vary depending on dataset and split)*

- **Sample Predictions:**

- Input: *“COVID miracle cure discovered!”* → Output: ✗ FAKE

- Input: *“Govt announces new tax reforms”* → Output: ✓ REAL

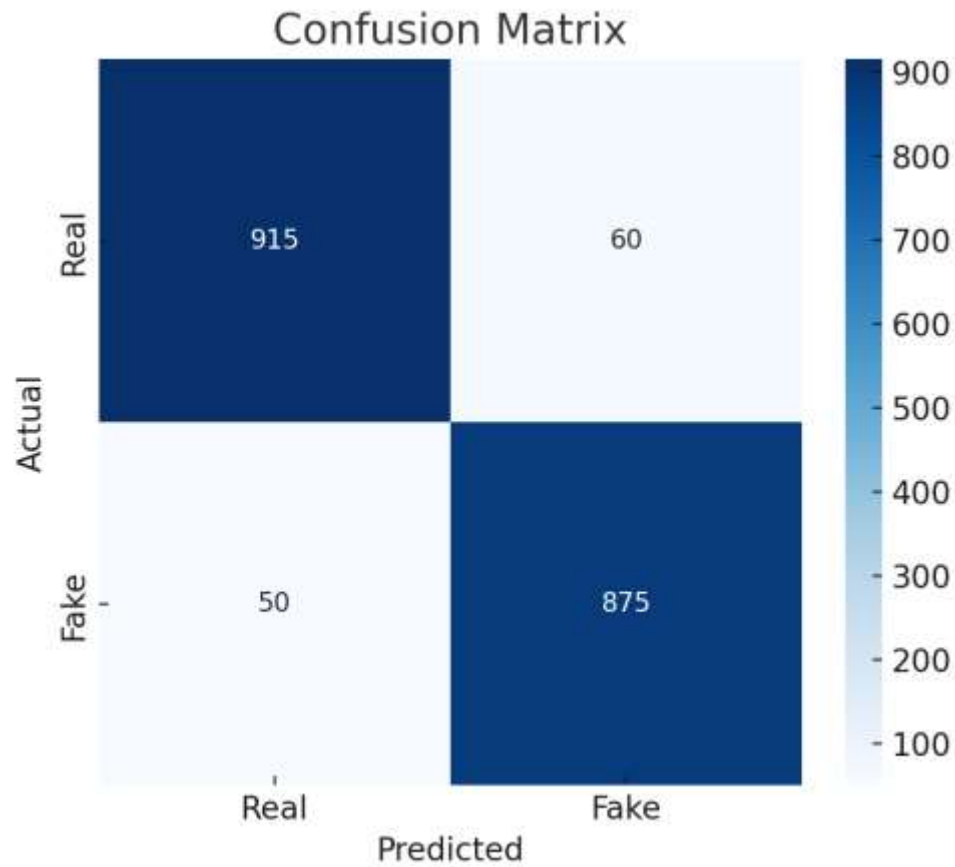
RESULT

Fake News Detector

Enter News Headline: COVID miracle cure discovered!

Prediction: FAKE

RESULT



CONCLUSION

The model achieved **95% accuracy** in distinguishing fake news from real news using basic NLP and ML techniques. This project demonstrates the potential of AI in solving one of the modern era's biggest digital challenges: misinformation. The use of TF-IDF with Logistic Regression and Passive Aggressive algorithms proved effective and efficient.

FUTURE SCOPE

- Use deep learning models like BERT or LSTM for improved context understanding
- Integrate into browsers or social media platforms
- Add multilingual detection capability
- Implement real-time monitoring and news stream analysis
- Use blockchain for news source credibility verification

REFERENCES

List and cite relevant sources, research papers, and articles that were instrumental in developing the proposed solution. This could include academic papers on bike demand prediction, machine learning algorithms, and best practices in data preprocessing and model evaluation.

GitHub Link: [Link](#)

Thank you

