



TRAC BUILDX

Problem Statement Title- AI-POWERED FAKE NEWS

DETECTOR

Team Name - TRUTH TITANS

Team Leader - MAYANK RAJ

Member 1- MAYANK RAJ

Member 2- JAGRITI CHHABRA

Member 3- UDIT TIWARI

PROPOSED SOLUTION



We propose a Machine Learning-based web application that automatically analyzes news articles and predicts whether they are Real or Fake, along with their authenticity score.

Step 1: User Input

User pastes:

News article text

OR article URL

System extracts the main content

Step 2: Text Preprocessing (NLP)

Remove stopwords

Convert to lowercase

Remove punctuation

Tokenization

Lemmatization

This cleans the text for accurate analysis.

BUILD X

Step 3: Feature Extraction

Use TF-IDF (Term Frequency–Inverse Document Frequency)

Converts text into numerical format

Highlights important words in the article

Step 4: Machine Learning Model

Logistic Regression (or Naive Bayes)

Trained on labeled dataset (Real = 0, Fake = 1)

Model learns writing patterns of fake news

Step 5: Prediction & Authenticity Score

System predicts:

✓ Real News

⚠ Fake News

Generates:

Authenticity Score (0–100%)

Fake Probability (%)

TECHNICAL APPROACH



1. Data Collection

Dataset collected from:

Kaggle Fake News Dataset

News APIs

Labeled data:

0 → Real

1 → Fake

Balanced dataset to avoid bias

2. Data Preprocessing (NLP Pipeline)

Text cleaning steps:

Lowercasing

Removing punctuation & special characters

Stopword removal

Tokenization

Lemmatization / Stemming

3. Feature Engineering

TF-IDF Vectorization

Converts text into numerical feature vectors

Captures importance of words

Reduces impact of common words

Formula:

$$\text{TF-IDF} = \text{TF} \times \log(\frac{N}{DF})$$

Where:

TF = Term Frequency

DF = Document Frequency

N = Total Documents

FEASIBILITY AND VIABILITY



Technical Feasibility

✓ The project uses proven technologies:

Python
Scikit-learn
NLP (TF-IDF)
Logistic Regression

✓ Requires:

Basic dataset (Kaggle available)
Standard laptop (8GB RAM enough)
No expensive hardware needed

✓ Model training time:

Few minutes on normal system

✓ Deployment possible via:

Streamlit
Flask
Web hosting platforms

Conclusion: Technically easy to implement within hackathon time.

Market Viability

✓ Huge demand due to:

Rise in misinformation
Social media fake news spread
Political propaganda
Health misinformation

✓ Target Users:

Students
Journalists
Media houses
Social media users
Fact-checking organizations

✓ Potential Applications:

Browser Extension
News verification platform
Educational tool

 **Conclusion:** Strong real-world relevance and scalability.

IMPACT AND BENEFITS



Social Impact:

- ✓ Reduces spread of misinformation
- ✓ Promotes responsible content sharing
- ✓ Encourages critical thinking
- ✓ Builds digital awareness among youth

↳ Helps create a more informed society.

Educational Impact:

- ✓ Helps students verify research sources
- ✓ Useful for academic projects
- ✓ Improves media literacy skills

↳ Makes students smarter digital citizens.

Public Safety Impact:

- ✓ Prevents panic caused by fake health news
- ✓ Reduces political misinformation
- ✓ Helps stop rumor-based violence

↳ Contributes to safer communities.

Technological Benefits

- ✓ Demonstrates practical use of AI & NLP
- ✓ Scalable to real-time systems

Can integrate with:

Browser extensions
News websites
Social media platforms
Shows real-world AI application.

Economic Benefits

- ✓ Low development cost
- ✓ Open-source tools

Can become:

SaaS platform
API service
Media verification tool
Strong startup potential 

RESEARCH AND REFERENCES



Kaggle (Data science datasets)

Research papers from Google Scholar

NLP tutorials and ML model comparisons

“Liar, Liar Pants on Fire”

Authors: William Yang Wang

Dataset: LIAR dataset

Contribution: Political fact-checking dataset for ML training

“Automatic Fake News Detection”

Focus: Deep learning-based fake news classification

Compared traditional ML vs Neural Networks

THANK YOU
BUILD X

by:TRUTH
TITANS