

Project Report: Fake News Detection System

Organization: TruthLens Institute

Author: Habib Bashir Lawal

Date: 06/28/2025

1. Executive Summary

Developed a machine learning system to detect fake news with 85% accuracy

Analyzed 12,999 articles from 244 domains

Key findings: Sensationalist language patterns, high-risk domains identified

Best model: Logistic Regression (F1-score: 0.85)

2. Methodology

Data Collection

Source: BS Detector extension dataset

12,999 records with text, URLs, and labels

3. Key Results:

Exploratory Findings

Top fake news domains:

python

['infowars.com', 'breitbart.com', 'naturalnews.com']

Most common fake news words:

Model Performance

Model	Accuracy	F1-Score
Logistic Regression	86%	0.85
Random Forest	83%	0.82

4. Challenges & Solutions

Challenge	Solution Applied
Class imbalance	Stratified train-test split
Noisy text data	Advanced NLP cleaning pipeline
Limited metadata	Extracted domains from URLs

5. Ethical Considerations

Risk of overflagging legitimate satire/news