

Fake News Detection Project Report

1. Title:

Fake News Detection Using Machine Learning

2. Introduction:

The rise of misinformation and fake news on the internet presents a significant challenge to societies worldwide. Detecting fake news automatically is essential to prevent the spread of false information and ensure public awareness is based on verified facts.

3. Problem Statement:

To develop a machine learning model that can classify news articles as **FAKE** or **REAL** based on the textual content.

4. Objectives:

- Preprocess text data from news articles.
- Apply feature extraction using TF-IDF.
- Train a machine learning classifier to distinguish between real and fake news.
- Save and deploy the model for future predictions.

5. Tools and Technologies Used:

- Python
- Pandas
- Scikit-learn
- TfidfVectorizer
- Logistic Regression
- Pickle (for model saving)

6. Dataset:

A small synthetic dataset was created containing news titles, text, and corresponding labels (REAL or FAKE). The dataset is stored in `fake_or_real_news.csv`.

7. Methodology:

Step 1: Data Preparation

- Created a DataFrame with sample news articles and labels.

- Saved data to a CSV file.

Step 2: Data Loading & Preprocessing

- Loaded the CSV file.
- Mapped labels: FAKE → 0, REAL → 1.

Step 3: Feature Extraction

- Applied **TF-IDF Vectorization** to convert text data into numerical vectors.

Step 4: Model Building

- Split data into training and testing sets.
- Used **Logistic Regression** for classification.

Step 5: Evaluation

- Achieved accuracy using `accuracy_score`.
- Evaluated results using `confusion_matrix`.

Step 6: Model Saving

- Trained model saved as `logistic_model.pkl` using pickle.

Step 7: Prediction Function

- Created a function `predict_news(text)` that predicts whether input text is FAKE or REAL.

8. Results:

- Model trained successfully.
- Accuracy score printed on test data.
- Function tested with example sentences.

9. Applications:

- Social media platforms to filter fake news.
- News agencies for content verification.
- Public information systems.

10. Limitations:

- Small dataset may limit real-world performance.
- May not capture nuances or sarcasm.

11. Future Work:

- Use a larger real-world dataset.
- Experiment with advanced models like Naive Bayes, SVM, or BERT.
- Build a user-friendly web interface.

12. Conclusion:

This project demonstrates a simple yet effective approach to detecting fake news using text-based machine learning techniques. The model provides a starting point for more sophisticated systems aimed at combating misinformation online.

13. References:

- Scikit-learn documentation
- Natural Language Processing research articles
- Kaggle datasets and tutorials

End of Report