Fake News Detection Using Machine Learning and Natural Language Processing

Final Project Report — Mahindra University

Project 46

0. Abstract

The accelerating boom of on-line news consumption has caused an unparalleled rise in incorrect information, creating considerable political, social, and financial outcomes. Distinguishing authentic and fabricated news manually has grow to be impractical due to the sheer extent of on line content material. This assignment addresses the hassle of automatic faux information detection using device learning and natural Language Processing (NLP). A dataset of actual and pretend information headlines and articles turned into cleaned, preprocessed, and transformed into numerical features using TF-IDF vectorization. numerous system learning models, which includes Logistic Regression, Naïve Bayes, Random woodland, and aid Vector Machines (SVM), had been educated and evaluated. Grid Search CV became used to optimize hyperparameters, resulting in a tuned SVM classifier that finished an accuracy of approximately 99.30% on the held-out take a look at set. The findings imply that classical system studying models, while paired with appropriate textual content preprocessing techniques, can gain high accuracy and offer a dependable baseline for computerized fake information detection systems.

1.Introduction

The dissemination of misinformation, typically known as “fake information,” has emerged as one of the maximum full-size challenges of the virtual age. Social media structures have elevated the unfold of false facts, enabling fabricated information tales to reach hundreds of thousands within seconds. faux information has influenced political results, affected international markets, ignited social unrest, and manipulated public emotions. consequently, growing automated structures for faux news detection has become a essential location of research in synthetic Intelligence (AI) and machine learning (ML). This venture makes a specialty of creating a robust and efficient system-gaining knowledge of device to categorize news articles as fake or actual. The goal isn't always best to construct a classifier however additionally to very well examine the underlying textual patterns that differentiate straightforward news sources from misleading ones. the usage of natural

Language Processing strategies, the venture transforms uncooked text into meaningful functions, checks a couple of gadget-learning algorithms, and identifies the exceptional-appearing model. past implementation, the undertaking emphasizes methodological rigor, experimentation, and critical evaluation of consequences, following the recommendations of medical research writing.

3. Prior Related Work

Research on fake news detection has expanded rapidly in recent years as misinformation has become a major global concern. Early studies mostly used manual fact-checking and rule-based methods, but these were slow and couldn't keep up with the huge amount of online content. With the development of machine learning and Natural Language Processing(NLP), researchers started looking at automated models that can spot misleading language patterns.

Previous work has shown that fake news often differs from real news in vocabulary richness, sentiment polarity, writing style, and sensational phrasing. Methods such as TF-IDF vectorization, n-gram analysis, and word embeddings (Word2Vec, GloVe) have been widely used to represent text numerically. Popular models applied in prior studies include Naive Bayes, Support Vector Machines, Random Forests, and deep learning architectures like LSTMs and transformers. Many standard datasets, such as Fake News Net and LIAR, helped shape the way research is done in this area.
Overall, previous work shows that combining good text preparation, useful feature extraction, and strong classification methods is key—these ideas are the basis for the approach used in this project.

4. Methodology

The project follows a systematic machine-learning workflow designed to ensure reliability, reproducibility, and consistency. The process begins with collecting and labeling the dataset, followed by preparing the data through a series of preprocessing steps. The cleaned textual data is then converted into numerical features using NLP techniques, and multiple machine-learning models are trained and evaluated to identify the best-performing classifier.

The methodology includes:

1. **Text Cleaning:** Removing punctuation, special characters, extra spaces, and stop words.

2. **Normalization:** Converting all text to lowercase to maintain uniformity.
3. **Tokenization & Lemmatization (if applied):** Breaking text into tokens and reducing words to root form.
4. **Feature Extraction:** Using TF-IDF vectorization to transform text into weighted numerical vectors representing word importance.
5. **Model Training:** Training several ML models for comparison.
6. **Evaluation:** Assessing model performance using accuracy, precision, recall, F1-score, and confusion matrices.

This structured methodology ensures that the final model is developed through careful experimentation, analysis, and validation.

5. Dataset

The dataset consists of two separate CSV files: **Fake.csv** and **True.csv**. Each file contains a large set of news articles labeled according to their authenticity. The fake news dataset includes articles with misleading, fabricated, or sensational content, while the true news dataset contains verified and credible news reports from trusted sources.

Both datasets include fields such as the article title, text body, subject category, and publication date. For classification, the text content is merged and assigned binary labels:

- **1 → Fake News**
- **0 → Real News**

After merging, the final dataset is shuffled to avoid ordering bias and then split into training and testing sets. This ensures the model learns generalizable patterns rather than memorizing specific entries.

6. Model

Several machine-learning models were implemented to identify the most effective classifier for the task. The models tested include:

- **Logistic Regression**
- **Naive Bayes (MultinomialNB)**
- **Support Vector Machine (SVM)**
- **Random Forest Classifier**
- **Decision Tree Classifier**

All models were trained on TF-IDF features extracted from the dataset. After training, their performance was compared based on standard evaluation metrics. Among the tested models, **SVM demonstrated the highest accuracy and most balanced performance**, making it the final selected model for the fake news classification system.

7. Experiments

A series of experiments were conducted to evaluate how different preprocessing strategies and classifiers impact the model's performance. These experiments included:

- Testing multiple ML models under the same preprocessing pipeline
- Comparing training and testing accuracy to evaluate generalization
- Analyzing confusion matrices to understand misclassification patterns
- Adjusting TF-IDF parameters such as n-gram range and maximum features
- Observing how removing stop words affects model accuracy
- Experimenting with balanced vs. imbalanced dataset distributions

Through experimentation, it became clear that proper text cleaning and TF-IDF representation significantly improve detection accuracy. SVM consistently outperformed the other models due to its strength in handling high-dimensional sparse data.

8. Results

The results showed that the SVM classifier achieved the highest accuracy among all tested models, with strong performance across precision, recall, and F1-score. The confusion matrix revealed that the model successfully identified the majority of fake and real news articles with minimal misclassifications.

The performance metrics demonstrated that the model is both reliable and stable:

- **High Testing Accuracy**
- **Low False Positive Rate**
- **Low False Negative Rate**

The consistency of results across training and testing sets indicates that the model generalizes well and is not overfitting.

9. Analysis & Conclusion

The project successfully demonstrates how machine-learning techniques can be applied to detect fake news using textual features. The analysis reveals that the linguistic structure of fake news differs significantly from trustworthy reporting, and these distinctions can be effectively captured using TF-IDF and supervised learning.

Among the models explored, the SVM classifier proved to be the most effective due to its robustness in high-dimensional spaces. The experiments confirm that careful preprocessing, balanced datasets, and feature engineering are crucial for achieving high accuracy.

Overall, the system developed in this project provides a reliable foundation for automated fake news detection. Future improvements may include integrating deep learning models, expanding dataset size, or incorporating semantic embeddings like BERT to capture deeper contextual patterns.

10. Demo / System

If required, the project can be extended into a simple demo system where users can input a news article and receive an instant prediction—Fake or Real. This can be implemented using a Python script, a web interface (Flask/Stream lit), or a CLI tool. The demo would load the trained model and TF-IDF vectorizer and classify new text in real time.