MACHINE LEARNING FOR TRUTH DISCOVERY: A REVIEW OF FAKE NEWS, PROPAGANDA AUTHORS, AND THEIR IMPACT

ABSTRACT

The proliferation of fake news and propaganda in the digital era has significantly influenced public opinion, political polarization, and societal trust. Identifying and mitigating misinformation has become a critical challenge for researchers, media institutions, and policymakers. This study presents a comprehensive analysis of machine learning-based truth discovery methods that focus on detecting, classifying, and preventing the dissemination of fabricated news content. The research emphasizes the integration of natural language processing (NLP), deep learning, and network-based models for the identification of propaganda authors and coordinated misinformation campaigns. The proposed review framework evaluates multiple approaches, including content-based linguistic analysis, user-behavior modeling, and source reliability estimation. Comparative analysis demonstrates that hybrid models combining deep neural networks with graph-based truth discovery outperform conventional classifiers in both accuracy and generalizability. The findings highlight the transformative potential of intelligent truth-verification systems in combating online misinformation and restoring digital credibility across global media platforms.

Keywords: Fake News, Truth Discovery, Machine Learning, Propaganda Detection, Author Profiling, Natural Language Processing.

EXISTING SYSTEM

Existing systems for fake news detection rely heavily on supervised machine learning and rule-based models that primarily focus on text classification. These models are typically trained on labeled datasets containing factual and non-factual samples. They often utilize shallow linguistic features such as word frequency, sentiment polarity, and n-gram statistics. While effective for simple patterns, these models struggle with the semantic complexity and contextual variations inherent in propaganda and misinformation. Moreover, they tend to perform poorly when applied to emerging events or new linguistic expressions not present in the training data.

In addition to text-based detection, network analysis techniques have been used to identify misinformation propagation by tracking retweet networks and user engagement metrics.

However, such approaches depend on extensive data availability and are vulnerable to manipulation by coordinated bot networks. Most existing frameworks also lack the ability to perform author profiling or identify stylistic markers of propaganda generation. As a result, they can detect misleading content but fail to trace its origins or intent.

Disadvantages of the Existing System

1. **Limited Contextual Understanding:**

Traditional models based on shallow linguistic features fail to capture complex semantics and deeper contextual relationships in news articles.

2. **Lack of Author Attribution:**

Existing frameworks detect fake content but rarely identify or analyze the authors responsible for generating propaganda or misinformation.

3. **Low Adaptability and Generalization:**

Many models trained on specific datasets fail to generalize across platforms, languages, or cultural contexts, reducing their effectiveness in real-world scenarios.

PROPOSED SYSTEM

The proposed framework leverages machine learning for holistic truth discovery, integrating deep learning, natural language processing, and network-based author attribution to identify and mitigate fake news and propaganda effectively. The model employs a hybrid architecture that combines contextual embeddings from transformer models with relational analysis through graph neural networks (GNNs). This integration allows simultaneous assessment of textual semantics, author behavior, and propagation networks, offering a multi-dimensional approach to truth verification.

At the content level, transformer-based language models such as BERT and RoBERTa capture deep contextual representations of text, enabling the detection of subtle deception cues and rhetorical techniques. These embeddings are fed into a truth discovery module that evaluates the factual consistency of content using external knowledge graphs. The author profiling component employs stylometric and behavioral analysis to uncover linguistic patterns, writing consistency, and topic bias across multiple publications. This facilitates accurate identification of propaganda authors and coordinated misinformation campaigns.

The system also incorporates a graph-based propagation analyzer that models the diffusion of information across social networks. By detecting clusters of synchronized accounts or repetitive

posting patterns, it becomes possible to uncover the hidden architecture of disinformation networks. The integration of all three modules—content analysis, author profiling, and propagation tracking—ensures comprehensive truth discovery capable of both detection and attribution.

Advantages of the Proposed System

1. **Comprehensive Truth Discovery:**

Combines content-level linguistic analysis, author profiling, and network propagation modeling for accurate misinformation detection and source tracing.

2. **Adaptive and Scalable Framework:**

Employs transformer and graph-based architectures that adapt to evolving linguistic trends, ensuring high accuracy across diverse languages and domains.

3. **Author Identification and Behavioral Analysis:**

Enables the detection of propaganda authors by analyzing writing style, topic preference, and dissemination behavior, contributing to transparent digital journalism.

SYSTEM REQUIREMENTS

> ➢ **H/W System Configuration:-**

> ➢ Processor   -   Pentium –IV

> ➢ RAM   - 4 GB (min)

> ➢ Hard Disk   -   20 GB

> ➢ Key Board   -   Standard Windows Keyboard

> ➢ Mouse   -   Two or Three Button Mouse

> ➢ Monitor   -   SVGA

SOFTWARE REQUIREMENTS:

> ❖ **Operating system**   : Windows 7 Ultimate.
> ❖ **Coding Language**   : Python.
> ❖ **Front-End**   : Python.
> ❖ **Back-End**   : Django-ORM
> ❖ **Designing**   : Html, css, javascript.
> ❖ **Data Base**   : MySQL (WAMP Server).