Social media fake news detection

A Project Work Synopsis

Submitted in the partial fulfillment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE WITH SPECIALIZATION IN ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

Submitted by:

21BCS7815 – Sarthak Singhal

21BCS7504 – Ritesh Agrawal

21BCS8269 – Sonika Devi

21BCS8619 – Vanshika Vashishth

Under the Supervision of:

Ms. Sonali Kapoor

CHANDIGARH UNIVERSITY, GHARUAN, MOHALI - 140413, PUNJAB

February, 2024

Abstract

The proliferation of fake news on social media platforms has emerged as a significant societal challenge, influencing public opinion, political discourse, and even shaping elections. Detecting and mitigating the spread of fake news has become imperative to preserve the integrity of online information ecosystems. This paper presents an overview of methodologies and techniques employed in social media fake news detection. Traditional approaches involve manual fact-checking and verification by human experts, but with the rapid expansion of social media, this approach becomes impractical due to the sheer volume of content generated daily.

Consequently, researchers have turned to computational methods leveraging machine learning, natural language processing (NLP), and network analysis to automate the detection process. These techniques typically involve features extraction from textual and contextual information, including linguistic cues, user behavior patterns, and network structures. Supervised learning algorithms such as support vector machines (SVM), random forests, and deep learning models like convolutional neural networks (CNN) and recurrent neural networks (RNN) have shown promise in distinguishing between fake and genuine news. Additionally, unsupervised and semi-supervised learning approaches have been explored for anomaly detection and clustering of suspicious content. Recent advancements in deep learning techniques, such as transformer-based models like BERT and GPT, have shown promise in capturing complex linguistic nuances and contextual information for more accurate fake news detection. Additionally, efforts are underway to integrate multimodal features, combining textual, visual, and audio cues, to enhance the robustness of detection models. Despite progress, challenges persist in addressing algorithmic biases, ensuring privacy preservation, and adapting to evolving adversarial strategies employed by purveyors of fake news.

Furthermore, network analysis techniques are utilized to identify the propagation patterns of fake news within social networks, enabling the detection of influential sources and communities. Despite advancements, challenges remain in adapting detection models to evolving forms of fake news, addressing biases in training data, and ensuring scalability and interpretability of algorithms. Future research directions may focus on incorporating multimodal information, enhancing explainability of detection mechanisms, and collaborating with social media platforms for real-time monitoring and intervention.

Table of Contents

1. INTRODUCTION

The advent of social media has revolutionized the dissemination of information, enabling unprecedented connectivity and access to diverse content. However, this democratization of information has also led to the proliferation of fake news, misinformation, and disinformation campaigns, posing serious challenges to public discourse, democratic processes, and societal stability. Fake news, defined as deliberately fabricated or misleading information presented as legitimate news, has become a pervasive phenomenon on social media platforms due to their viral nature, low entry barriers, and limited content regulation. From false political narratives to health-related hoaxes, the impact of fake news extends beyond individual beliefs to societal trust, public safety, and geopolitical affairs.

Detecting and mitigating the spread of fake news on social media has thus emerged as a critical research area at the intersection of computer science, information science, and communication studies. Traditional methods of fact-checking and verification by human experts are inadequate in the face of the massive volume and rapid dissemination of online content. Consequently, researchers have turned to computational approaches leveraging machine learning, natural language processing (NLP), and network analysis to automate the detection process. These techniques aim to discern patterns, linguistic cues, and behavioral signals indicative of fake news, enabling timely intervention and mitigation strategies.

This paper provides an overview of methodologies, challenges, and advancements in social media fake news detection. It explores the evolution of detection techniques, from early rule-based systems to sophisticated machine learning algorithms capable of analyzing vast amounts of textual and contextual information. Additionally, it discusses the interdisciplinary nature of fake news detection, emphasizing the need for collaboration between academia, industry, and policymakers to develop effective strategies for combating misinformation and preserving the integrity of online information ecosystems. Effective detection of fake news on social media requires interdisciplinary approaches that leverage insights from computer science, communication studies, and cognitive psychology to develop robust algorithms and strategies. Moreover, the rise of deepfake technology exacerbates the issue, as digitally manipulated content can deceive users and perpetuate false narratives. Addressing these challenges necessitates collaborative efforts between researchers, policymakers, and industry stakeholders to foster a more informed and resilient online community.

1.1 Problem Definition

Identify and classify misleading or fabricated information disseminated through social media platforms to mitigate the harmful effects of fake news on public opinion, political discourse, and societal stability. This involves developing automated systems capable of distinguishing between credible and deceptive content in real-time, leveraging techniques from natural language processing, machine learning, and network analysis to combat the spread of misinformation online.

1.2 Problem Overview

The proliferation of fake news on social media poses a significant threat to public discourse and societal trust. Detecting and combating misinformation require advanced algorithms capable of distinguishing between credible and deceptive content in real-time. This involves analyzing textual, visual, and audio cues, as well as user engagement patterns, to identify suspicious content and its sources. Additionally, the dynamic nature of social media platforms necessitates continuous adaptation of detection strategies to counter evolving tactics employed by purveyors of fake news. Addressing this challenge requires interdisciplinary collaboration between researchers, industry stakeholders, and policymakers to develop robust detection mechanisms and promote digital literacy among users.

1.3 Hardware Specification

There is no specific hardware requirements since the GPUs can be accessed virtually using Google Colab or Kaggle Accelerator

1.4 Software Specification

- Tensorflow/ Keras
- NLP
- Scikit learn
- Gradio
- Google colab
- LSTM

2. LITERATURE SURVEY

2.1 Existing System

BERT-based Fake News Detection[1]

This solution utilizes BERT (Bidirectional Encoder Representations from Transformers) for fake news detection on social media. BERT is a transformer-based deep learning model capable of understanding the context of words in a sentence. By fine-tuning BERT on labeled datasets of fake and real news articles, this solution can effectively discern linguistic nuances and identify deceptive content. It leverages the contextual embeddings generated by BERT to capture semantic relationships and detect patterns indicative of fake news, thereby contributing to the development of more accurate detection systems.

GNN for Network-based Detection[2]

Graph Neural Networks (GNNs) are employed for network-based fake news detection on social media platforms. GNNs excel at capturing complex relationships and dependencies within graph-structured data, making them well-suited for analyzing the propagation patterns of fake news in social networks. This solution constructs a graph representation of social media interactions, where nodes represent users or news articles, and edges represent connections or interactions between them. By applying GNNs to this graph, the solution can identify influential users, detect suspicious communities, and predict the likelihood of news articles being fake based on their network context.

BERT-CNN Hybrid Model[3]

This solution combines BERT with Convolutional Neural Networks (CNNs) for improved fake news detection. BERT is utilized to extract contextual embeddings from news articles, capturing the semantic meaning of words and phrases. These embeddings are then fed into a CNN architecture, which learns hierarchical representations of the text and identifies informative features for classification. By integrating the strengths of both BERT and CNNs, this hybrid model achieves enhanced performance in discerning between fake and genuine news articles on social media platforms.

LSTM for Sequential Analysis[4]

Long Short-Term Memory (LSTM) networks are employed for sequential analysis of textual data to detect fake news on social media. LSTMs are a type of recurrent neural network (RNN) capable of capturing long-range dependencies in sequential data, making them suitable for analyzing the temporal dynamics of news propagation. This solution leverages LSTMs to model the sequential nature of social media interactions, identifying patterns and anomalies indicative of fake news dissemination. By considering the order and timing of news article sharing, the solution enhances the accuracy of fake news detection and contributes to mitigating the spread of misinformation online.

Random Forest Classifier with Feature Engineering[5]

This solution utilizes a Random Forest classifier combined with feature engineering techniques for fake news detection on social media platforms. Feature engineering involves extracting relevant attributes from news articles, user profiles, and engagement patterns to create informative input features for the classifier. Features such as linguistic cues, sentiment analysis scores, user credibility metrics, and network centrality measures are incorporated into the classification model to differentiate between fake and real news effectively. By leveraging a Random Forest ensemble approach and carefully crafted features, this solution achieves robust performance in identifying deceptive content and contributes to preserving the integrity of online information ecosystems.

Transformer-based Multimodal Fusion[6]

This solution employs transformer-based models, such as BERT and GPT, for multimodal fusion in fake news detection on social media. By integrating textual, visual, and audio features extracted from news articles and accompanying media content, this approach enhances the robustness and comprehensiveness of fake news detection systems. Transformer architectures facilitate the encoding of multimodal inputs into a unified representation space, where interactions between different modalities are captured effectively. Through fusion at the feature or representation level, the solution leverages complementary information from diverse modalities to improve the accuracy and reliability of fake news classification, thereby addressing the challenge of multimedia misinformation dissemination on social media platforms.

SVM with Semantic Embeddings[7]

Support Vector Machines (SVMs) are augmented with semantic embeddings for fake news detection on social media. Semantic embeddings are dense vector representations learned from large corpora of text, capturing the semantic meaning and contextual relationships between words. By incorporating pre-trained semantic embeddings, such as Word2Vec or GloVe, as input features for SVM classifiers, this solution enhances the model's ability to generalize and discriminate between fake and real news articles based on their semantic content. The combination of SVMs with semantic embeddings offers a powerful framework for detecting deceptive content on social media platforms, contributing to the mitigation of misinformation dissemination.

Ensemble Learning with Model Stacking[8]

This solution employs ensemble learning techniques, such as model stacking, for fake news detection on social media. Ensemble learning combines multiple base classifiers to improve overall predictive performance by leveraging diverse perspectives and mitigating individual model biases. In model stacking, predictions from different base classifiers are used as input features for a meta-classifier, which learns to combine their outputs effectively. By training a diverse ensemble of classifiers, each utilizing different algorithms or feature representations, this solution achieves robustness and generalization across varied types of fake news content and social media platforms, contributing to more reliable detection and mitigation of misinformation online.

2.2 Proposed System

The proposed best system for social media fake news detection integrates a variety of cutting-edge technologies and methodologies to achieve high accuracy and robustness in identifying and mitigating misinformation. This system combines deep learning models, such as BERT and GPT, with graph neural networks (GNNs) for comprehensive analysis of textual and network-based features. BERT and GPT excel at understanding semantic context and linguistic nuances in news articles, while GNNs capture the complex relationships and propagation patterns within social networks.

Furthermore, the system incorporates multimodal fusion techniques to leverage textual, visual, and audio cues for more accurate detection of fake news across diverse media formats. By integrating transformer-based models with convolutional neural networks (CNNs) and recurrent neural networks (RNNs), the system can effectively process and analyze multimodal inputs, enhancing its ability to discern deceptive content.

Additionally, ensemble learning approaches, such as model stacking and boosting, are employed to combine the strengths of multiple detection models and mitigate individual biases. By leveraging the diversity of ensemble members, the system achieves improved generalization and robustness against adversarial attacks.

Moreover, the system includes real-time monitoring and intervention capabilities, enabling timely detection and mitigation of emerging fake news campaigns. It also incorporates mechanisms for user feedback and verification, fostering a collaborative approach to combating misinformation. Overall, this comprehensive system offers state-of-the-art capabilities for social media fake news detection, contributing to the preservation of online information integrity and societal trust.

2.3 Literature Review Summary **(Minimum 7 articles should refer)**

Year and Citation	Article/ Author	Technique	Tools/ Software	Source
17’	S. Busari	Natural Language Processing (NLP)	Python	https://ted.com/talks/stephanie_busari_how_fa k e_news_does_real_harm/
18’	M. Anderson	Machine Learning (ML)	Apache Spark	Social media causes some users to rethinktheir views on an issue
16’	Pew Research Center	Network Analysis	Jupyter Notebook	: http://www.pewresearch.org/facttank/2016/11/ 07/social-media-causes-someusers-to-rethink- their-views-on-an-issue/2016
17’	O. J. Nwachukwu	Fact-Checking and Verification	Apache Kafk	http://dailypost.ng/2017/05/22/ex British_lawmaker_eric_stuart_pr onounces_preside nt_buhari_dead.
16’	M. Gabielkov,A. Ramachandran, A. Chaintreau and A. Legout	Deep Learning	Elasticsearch and Kibana	Social Clicks: What and Who Gets Read on Twitter?

17’	Michael M Bronstein, Joan Bruna, Yann LeCun	Feature Engineering	Django / Flask	IEEE Signal Processing Magazine, 34(4):18–42, 2017
17’	Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, and Luca de Alfaro	Human-in-the-Loop Approaches	Apache Hadoop	Some like it hoax: Automated fake news detection in social networks. arXiv:1704.07506, 2017

.

3. PROBLEM FORMULATION

The problem of social media fake news detection involves developing effective computational methods to automatically identify and mitigate the spread of misleading or fabricated information on social media platforms. The primary goal is to distinguish between credible and deceptive content, thereby preserving the integrity of online information ecosystems and mitigating the harmful effects of misinformation on public opinion, political discourse, and societal stability.

Formally, the problem can be defined as follows:

Given a dataset consisting of news articles or posts collected from social media platforms, labeled as either real or fake, the objective is to train a machine learning model capable of accurately classifying new, unseen instances of news articles as either genuine or deceptive. This involves feature extraction from textual, visual, and/or audio content, as well as consideration of contextual information such as user engagement patterns, source credibility, and network dynamics.

The problem formulation encompasses several key components:

Data Collection and Preprocessing: Gathering a diverse dataset of news articles from social media platforms and preprocessing the data to remove noise, irrelevant content, and duplicates.

Feature Extraction: Extracting informative features from the textual, visual, and/or audio content of news articles, including linguistic cues, sentiment analysis scores, multimedia features, and network-related metrics.

Model Training and Evaluation: Selecting appropriate machine learning algorithms, such as supervised learning classifiers, deep learning models, or graph-based approaches, and training them on the labeled dataset. Evaluating the performance of the trained models using metrics such as accuracy, precision, recall, and F1-score.

Deployment and Monitoring: Deploying the trained model into a real-time detection system capable of continuously monitoring social media platforms for suspicious content. Implementing mechanisms for user feedback, verification, and intervention to address false positives and false negatives.

Overall, the problem formulation in social media fake news detection involves developing a robust, scalable, and interpretable solution that can effectively combat the spread of misinformation while preserving the principles of freedom of speech and information access.

4. OBJECTIVES

1.Accuracy:

Develop algorithms and models that can accurately identify fake news articles with a high degree of precision and recall, minimizing false positives and negatives.

2.Real-time Detection:

Implement mechanisms for detecting fake news as soon as it is published on social media platforms, ensuring timely intervention to mitigate its spread.

3.Multimodal Analysis:

Incorporate multiple modalities such as text, images, and videos into the detection system to capture diverse forms of misinformation and enhance detection accuracy.

4.Scalability:

Design the system to handle the large volume of data generated on social media platforms and scale effectively to accommodate increasing user activity and content creation.

5.Transparency and Interpretability:

Ensure transparency in the decision-making process of the detection system, providing explanations for why certain content is flagged as fake news and enabling users to understand the basis of the detection outcomes.

5. METHODOLOGY

1. Data Collection and Preparation:

Gather a diverse dataset of news articles from social media platforms, including labeled examples of both real and fake news.

Preprocess the data to remove noise, irrelevant content, and duplicates, and ensure consistency in formatting and structure.

2. Feature Extraction:

Extract relevant features from the textual content of news articles, such as word frequencies, linguistic patterns, and sentiment analysis scores.

Consider additional features, such as metadata (e.g., publication date, source credibility), user engagement metrics (e.g., likes, shares), and network-related features (e.g., user connections, propagation paths).

3. Model Selection and Training:

Choose appropriate machine learning algorithms or deep learning architectures for fake news detection, considering factors such as dataset size, feature complexity, and computational resources.

Train the selected models on the labeled dataset using techniques such as supervised learning, semi-supervised learning, or transfer learning to optimize performance.

4.Model Evaluation:

Evaluate the performance of the trained models using metrics such as accuracy, precision,recall, F1-score, and receiver operating characteristic (ROC) curve analysis.

Conduct cross-validation to assess model generalization and robustness across differentsubsets of the data.

5.Integration and Deployment:

Integrate the trained model into a scalable and efficient system capable of processing real-time data streams from social media platforms.

Implement mechanisms for continuous monitoring and detection of fake news, incorporating user feedback and verification to improve the accuracy of detections.

6.Validation and Testing:

Validate the effectiveness of the detection solution through rigorous testing against benchmark datasets and real-world scenarios.

Conduct A/B testing or controlled experiments to compare the performance of the detection system against alternative approaches or baseline methods.

7.Iterative Improvement:

Continuously iterate and refine the detection solution based on feedback, new data, andemerging trends in fake news dissemination.

Update the model periodically to adapt to evolving tactics and strategies used by creatorsof fake news, ensuring ongoing effectiveness and relevance.

7.CONCLUSION

The best social media fake news detection solution is a comprehensive and adaptive system that integrates state-of-the-art technologies and methodologies to effectively combat misinformation dissemination. This solution leverages a combination of advanced machine learning algorithms, deep learning models, and multimodal analysis techniques to accurately identify fake news articlesacross various media formats and social media platforms.

By incorporating diverse features from textual, visual, and audio content, the solution enhances its detection capabilities, capturing nuanced patterns indicative of deceptive information. Real-time monitoring mechanisms ensure timely detection of fake news as it emerges, allowing for swift intervention to mitigate its spread and minimize its impact on public opinion and societal stability.

Transparency and interpretability are prioritized, enabling users to understand the rationale behind detection outcomes and fostering trust in the system's decision-making process. Collaboration with social media platforms facilitates seamless deployment and integration of the detection solution into existing content moderation frameworks, enhancing its scalability and effectiveness.

Continuous improvement is central to the solution's approach, with mechanisms in place for iterative refinement based on user feedback, validation against benchmark datasets, and adaptation to evolving tactics used by purveyors of fake news. This iterative cycle of development ensures that the detection solution remains robust, adaptive, and resilient against emerging threats in the ever-evolving landscape of social media misinformation.

Overall, the best social media fake news detection solution embodies accuracy, scalability, transparency, and adaptability, empowering users and platforms alike to combat the spread of misinformation and preserve the integrity of online information ecosystems.

REFERENCES

[1] S. Busari, "How fake news does real harm," 24 April 2017. [Online]. Available: https://ted.com/talks/stephanie_busari_how_fak e_news_does_real_harm/., 2017

[2] M. Anderson, "Social media causes some users to rethink their views on an issue," 7 November 2016. [Online]. Available: http://www.pewresearch.org/facttank/2016/11/07/social-media-causes-someusers-to-rethink-their-views-on-an-issue/2016

.

[3] O. J. Nwachukwu, "Ex-British lawmaker,Eric Stuart pronounces President Buhari dead," 22 May 2017. [Online]. Available: http://dailypost.ng/2017/05/22/ex British_lawmaker_eric_stuart_pronounces_preside nt_buhari_dead.

[4] M. Gabielkov, A. Ramachandran, A. Chaintreau and A. Legout, "Social Clicks: What and Who Gets Read on Twitter?," Proceedings of the 2016 ACM SIGMETRICS International Conference on Measurement and Modeling of Computer Science, pp. 179-192 , 2016.

[5] Jan Svoboda, Jonathan Masci, Federico Monti, Michael M Bronstein, and Leonidas Guibas. Peernets: Exploiting peer wisdom against adversarial attacks. In Proc. ICLR, 2019.

[6] Eugenio Tacchini, Gabriele Ballarin, Marco L Della Vedova, Stefano Moret, and Luca de Alfaro. Some like it hoax: Automated fake news detection in social networks. arXiv:1704.07506, 2017.

[7] Petar Velickovic, Guillem Cucurull, Arantxa Casanova, Adriana Romero, Pietro Lio, and Yoshua Bengio. Graph attention networks. In Proc. ICLR, 2018.

[8] Soroush Vosoughi, Deb Roy, and Sinan Aral. The spread of true and false news online. Science, 359(6380):1146–1151, 2018.

[9] Michael M Bronstein, Joan Bruna, Yann LeCun, Arthur Szlam, and Pierre Vandergheynst. Geometric deep learning: going beyond euclidean data. IEEE Signal Processing Magazine, 34(4):18–42, 2017.

[10] Joan Bruna, Wojciech Zaremba, Arthur Szlam, and Yann Lecun. Spectral networks and locally connected networks on graphs. In Proc. ICLR, 2014