Chapter 10
Combating Misinformation in Social Media and News:
A Deep Learning Approach

Usharani Bhimavarapu
https://orcid.org/0000-0002-0246-1420
Department of Computer Science and Engineering, Koneru Lakshmaiah Education Foundation, Vaddeswaram, India

ABSTRACT

Combating misinformation has become a critical challenge in today's information-driven society, particularly with the proliferation of fake news, propaganda, and biased content across various domains. This study explores advanced natural language processing (NLP) techniques, including feature extraction and selection, to analyze and classify datasets such as Q-Prop, ISOT, GRAFN, and PubHealth. The relief algorithm is employed for feature selection to identify the most relevant attributes, enhancing the efficiency of machine learning models. XLNet, a powerful transformer-based model, is utilized for document representation and classification due to its ability to capture bidirectional and long-term contextual dependencies. The proposed methodology demonstrates how robust embeddings, combined with domain-specific datasets and optimized feature selection, can accurately classify content across news, politics, and public health domains.

DOI: 10.4018/979-8-3693-9015-3.ch010

INTRODUCTION

Social media has transformed the way people consume and engage with news, reshaping the global media landscape. Platforms like Twitter, Facebook, Instagram, and YouTube have become primary sources of news dissemination, allowing real-time updates and instant access to information. The interactive nature of social media enables users to share, comment, and discuss news, fostering a participatory approach to information consumption. This shift has democratized news creation and distribution, empowering individuals and smaller organizations to share their perspectives alongside traditional media outlets.

One of the defining characteristics of social media and news is the speed at which information spreads. Breaking news stories can go viral within minutes, reaching millions of users worldwide. This immediacy, however, comes with challenges, such as the difficulty of verifying information, leading to the proliferation of misinformation and fake news. The algorithms that drive social media platforms often prioritize engagement over accuracy, amplifying sensational or polarizing content that may not always align with journalistic integrity.

Social media has also played a pivotal role in citizen journalism, where individuals on the ground document events as they unfold. This has proven invaluable during crises, protests, and natural disasters, offering firsthand accounts and visual evidence. Yet, the lack of editorial oversight raises concerns about biases, incomplete narratives, and ethical considerations in reporting.

Additionally, social media's ability to personalize content through algorithms creates echo chambers, where users are exposed primarily to information that aligns with their views. This can deepen societal polarization, as individuals become less likely to encounter diverse perspectives. Despite these challenges, social media has also been a tool for raising awareness about critical issues, mobilizing social movements, and giving voice to marginalized communities.

The relationship between social media and news is dynamic and multifaceted. While it has expanded access to information and fostered global connectivity, it has also necessitated greater efforts to ensure accuracy, combat misinformation, and promote media literacy. As the boundaries between traditional journalism and social media blur, finding a balance between speed, engagement, and reliability remains a critical goal in the evolving digital age.

Naive Bayes Classifier assumes feature independence, which is unrealistic in natural language contexts, leading to reduced accuracy (Afchar et al., 2018). Support Vector Machines (SVM) struggle with large datasets and non-linear decision boundaries unless kernel methods are used, which increase computational complexity (Ahmed et al., 2018). Random Forest often overfits when the dataset is noisy or imbalanced and lacks interpretability for real-time decision-making (Barrón-Cedeño et

Towards a Possibilistic Information Retrieval System Using Semantic Query Expansion