Air journal of Artificial Intelligence

ISSN: 3106-7743 | 3106-7735
Volume 3, Issue 1, 2026

COMPARATIVE ANALYSIS OF MACHINE LEARNING MODELS FOR FAKE NEWS IDENTIFICATION AND CLASSIFICATION

Faisal Shahzad[*1], **Nazia Azim**[2]

[*1,2] *PhD Scholar, Department of Computer Science, Abdul Wali Khan University Mardan*

[*1]faisalshahzad.23@yahoo.com, [2]naziaazim31@gmail.com

Keywords
Social media, Fake News, Spreading of news, Symmetric analysis, Ontology, Machine Learning.

Article History
Received: 03 January 2026
Accepted: 15 March 2026
Published: 31 March 2026

Corresponding Author: *
Faisal Shahzad

Abstract

With the rapid growth of social media, nearly 70% of people now rely on these platforms as their primary news source. However, social media has also become a major channel for spreading misinformation and fake news. This paper introduces a semantic-based approach for fake news identification, designed to capture the depth of misinformation and enable dynamic decision-making. The proposed Fake News Identification Ontology (FNIOnt) categorizes news content into fictitious classes and performs semantic analysis on dataset content. The extracted features are evaluated using three machine learning classifiers—Random Forest (RF), Logistic Regression (LR), and Long Short-Term Memory (LSTM). Experimental results demonstrate that the proposed framework outperforms existing methods, achieving a 99% accuracy rate.

Despite these promising results, the adaptability of fake news creators presents an ongoing challenge. Techniques such as half-truths, subtle linguistic manipulation, and evolving writing styles make detection increasingly complex. Therefore, machine learning systems must evolve dynamically, learning from new data and adapting to emerging patterns of deception. Deep learning models—including Recurrent Neural Networks (RNNs), Convolutional Neural Networks (CNNs), and Transformers—further enhance detection by automatically capturing semantic and syntactic relationships in unstructured text. Unlike traditional models that rely on manual feature engineering, these neural approaches enable context- and sentiment-aware classification, strengthening robustness against nuanced manipulations.

This study underscores the potential of combining semantic analysis with advanced machine learning and deep learning methods to build adaptive, accurate, and scalable systems for fake news detection.

INTRODUCTION

On social media, fake news and information trigger distress and anxiety in the society as society consumed more and more news on its digital sources. The social media has the most extensive basis of information dissemination whereby 60 to 70 percent of the individuals are consuming information as a means to gain profit. Nevertheless, such tools as Facebook, Twitter, Instagram or Sina Weibo do not have detection and prevention mechanisms of rumors. The role of the researcher is to devise methods of detecting and classifying the unhealthy information being presented by genuine news and make rumor recognition and obsession a bitter concern Rumor identification and fixation among the news has become a burning issue in society today, and it spurs anxiety, panic, and stress out of many people. Since individuals are becoming more and more dependent on online sources of news and, particularly social media, it is possible to find that false or incorrect information spreads very quickly that can dramatically affect people and their communities. Such platforms are created to enable rapid transfers of information, and in most cases,

ISSN: 3106-7743 | 3106-7735
Volume 3, Issue 1, 2026

there are no applicable means to distinguish factual news and fake stories. This has been a failed effort as it has facilitated the spread of dangerous information that has the capability to stampede mass opinion, alter the truth and even cause mass disorders. The social media is the greatest and most powerful medium of information dissemination and the research indicates that 60-70 percent of individuals actually consume news to their personal advantage, be it just supporting their own perception, achieving emotional satisfaction or even belonging to the most current and trending issue. This type of a setting predisposes likelihood of proliferating sensationalistic or biased information and with the rising dominance of such platforms, the effects of fake news would increase as well. The speed of disseminating false information is higher than the time that it takes to confirm facts and the information in most cases reaches millions of people before the truth is established. Consequently, social media tends to be criticized more and more not only about its contribution to spreading misinformation, which can lead to significant social consequences, such as ideological polarization and endangering the health of the population, but also to economic damage [1]. Deep learning algorithms become more and more used in the sphere of classification since they are able to interpret and analyze big amounts of information in various spheres, including pictures, text, and sound. The current increased usage of deep learning methods is based on their daunting potential to discover intricate patterns and representations of raw and unstructured information data without having to produce the extraction of such features manually. In contrast to the obsolete machine learning algorithms, which may demand the cautious, or otherwise intricate, choosing and processing of the input features, the models of deep learning and the neural networks, in particular, can detect the trends in data by their own means, which enables their use in such activities as image recognition, language interpretation, and voice recognition, to stunning effect. Such automatic attribute extraction ability appears especially handy when processing large and complex data, as in the case of detecting misinformation.

Deep learning models will have a big benefit compared to traditional machine learning machines in case of fake news detection and classification. The inherent characteristics of fake news are the use of humor (tech), subtle language twisting plus sensationalism and emotive content and reference misalignments. Such deceptive patterns call upon advanced models that would evaluate not only an individual word or phrase but also the context and relationship between them. In specific, deep learning models, e.g. convolutional neural networks (CNNs) to interpret texts, or recurrent neural networks (RNNs), in particular long short-term memory (LSTM) models, are specially designed to understand such intricacies in sequences of words or sentences. This ability renders them more appropriate in reading textual data where the situation is paramount in distinguishing between factual and fake news. Moreover, deep neural networks work well with large amounts of data, which is a crucial characteristic in the fake news detection field since newly written articles are actively being published all over the world constantly. As the example, deep learning algorithms are able to process a big dataset, created of million of articles and with a possibility to extract small features in each article to find similarities that can say about a particular news, fabricated, misleading, or truthful. Similarly, there is also a possibility of calibrating these models to identify various kinds of fake news, which includes untrue stories, bias coverage, or deceptive headings, as can be found in the current media. Also, the capability of deep learning to capture multimodal information like integration of text, images, and even videos in one is another factor that can make systems of fake news detection even more effective. Several fake news contents also contain multimedia sources like pictures, videos, or any type of media as evidence of their argument. With the ability to combine image and text recognition and analysis, deep learning algorithms can give a more comprehensive picture of what has been classified potentially increasing the accuracy of the system. As an example we may consider a picture in an article which may be manipulated or depicted differently and deep learning models will be able to locate such anomalies combined with perusing the text content. Such algorithms, in particular, deep learning models, have revolutionized the detection of fake news by eliminating the need to extract features manually which is being traditionally one of the most time-consuming and time-consuming components of any

ISSN: 3106-7743 | 3106-7735
Volume 3, Issue 1, 2026

machine learning task. Deep learning techniques such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) have the advantage over human specialists, in that they can automatically discover relevant features in raw data of any type, such as text, images, or other media. Not only this saves a lot of time and resources but also increases the ability of the model to find complex patterns in the data that a human eye might not discern. Such sophisticated algorithms are especially effective in relating to fake news since they can detect more complex characteristics based on a language and context that are suggestive of covert or disingenuous behavior, e.g., sensationalist language, emotional engagement, or inconsistencies in the story. That is why CNNs and RNNs have become one of the best-recommended methods to identify fake news since they obtain abilities to improve themselves automatically due to consuming more data.

Specifically, more hybrid alternatives with the strengths of CNN and RNN models have proven tremendously practical in the task of fake news classifications. CNNs work well on spatial patterns and therefore may be used to find local structure in text, word n-grams or syntax dependencies common in false or deceptive news. RNNs, and in particular Long Short-Term Memory (LSTM) networks, on the other hand, are meant to capture dependencies between different elements of a sequence of data, which is essential in terms of flowing information and context in a news article. Hybrid network allows CNN and RNNs to work simultaneously to identify both global and local features of text therefore resulting in improved and robust fake news classification. Such hybrid filters can distinguish fake news articles and actual news articles based on both the specific words or phrases that are used as well as the general pattern or flow of the story which aids in identifying deceptive information over the actual one. [3]. Identification and classification of fake news through machine learning represents a young field that seeks to resolve the increasing problem of fake news online. Such methods usually depend on different measures that help properly identify and differentiate between fake news and the good ones. Content-based mechanisms are one of the popular methods: this one evaluates the news item itself, meaning, it verifies its authenticity. All these mechanisms make use of different elements that include linguistics pattern, text structures, sentiment analysis, and keyword detection to detect inconsistencies or suspicious things in the content. Machine learning models may find the possible rumors or phony stories comparing such properties to reliable resources or historical data.

The other strategy concerns the utilization of user features where the profile and pattern of behavior of the person sharing the news under consideration is considered. As an example, trends such as the past activity of the user, their interaction levels, and their levels of trustworthiness in a given community can play highly instrumental roles in determining whether a news story would be more likely to be true or fake. This approach takes into consideration that some user actions, like posting sensationalised or biased information frequently, can be indicative of misinformation spreading, [4] and Hybrid-based mechanisms of fake news detection and classification will mix and match several approaches, especially both the content and social context-based approaches, to harness the power of each approache separately and achieve a more thorough and correct identification of misinformation. Through this combination of the various depths of examination, the hybrid models will be able to evaluate the news articles not just based on what they say but also how they are shared across the social networks providing a complete look at the validity of the information. Content-based mechanisms on the other hand in the hybrid models center around the linguistic aspect, structure and semantics of a particular news article. It comprises methods, such as sentiment analysis, topic modeling, or detection of particular linguistic features or hooks, common in fake news, such as sensationalized language, emotive words, or misleading headlines. Background aspects, including the length of the article, presence of hyperlinks and the text format too, are seen to be determined with regard to the credibility of the content. Also there are temporal features that are considered to determine the speed at which the news circulates or whether the news changes with time as well. In the case of the distribution of fake news, there can also be some temporal patterns observable with the immediate spread of the publication or its repetition, which in turn is used to signify a misinformation campaign. [5].

The number of methods that integrate user stance into the normally conducted news categorizing is few, but the Stance-aware Reinforcement Learning Framework (SRLF) has been designed to address the challenge of labeling accuracy. Mrowca et al. (2017) examined the headlines to tag the positions of the news articles, whereas Giasemidis et al. (2018) calculated the position of the message to classify tweets with an implication of whether their content favors or denies the rumor at hand. Moreover, these developments help to realize the significance of the user perspectives in the situation of misinformation. Using user stance into the classification front, researchers aspire to enhance robustness of models, their capacity in identifying the subtle views and lastly the efficiency of misinformation detection strategies in social media and news sites [6]. The fake news has brought massive damages and disorientation to society, and its solution should rely on authentication and fidelity. Artificial neural networks such as RNN and VAE have been introduced to check news authenticity as well as identify user positions. Wang et al. (2017) identified users who had a higher likelihood of the diffusion of false news on streaming social media. Du et al. (2017) applied targeted information extractions of neural networks and LSTMs with TAN. The authors of the article by Chen et al. (2016) rely on rumors which are individual user posting styles and behavior to identify fake news [7].

Malhotra et al. (2020) used a synergic approach to establish verbal clues in original tweets, obtain semantic knowledge, and examine the structure of the spread. They employed the Graph Neural Network, comprising of Recurrent Neural Network to classify the news as being genuine or fake. Propagation path accuracy of this hybrid method of 0.86 exceeded accuracy of methods based only on content and social factors. This progression shows that there is effectiveness in putting in place several strategies to increase the news classification on a greater scale. With the combination of the advantages of the IA and NN, scientists are opening the path towards more significant and consistent approaches to fighting misinformation and enhancing overall accuracy of the news verification system. These methodologies may be instrumental in terms of giving users access to credible information in an environment that is increasingly becoming a complex digital world. [8]. With the use of social media, the possibility of sharing fake news is enhanced. Detection of misinformation needs the understanding of the mechanics of rumors spreading, definition of people most liable to engage in the nondescript information, and an understanding of what forms of misinformation are more likely to spread. The traditional approaches to rumors detection usually involve the user profiles, content-related analysis, and propagation channels. Nevertheless, the validity of the information is at times determined by the relationships a user holds including his followers and friends. This knowledge underscores the complexity of misinformation patterns in the social media settings. After focusing on the content as well as the social networks within which information moves, the researchers will be able to create more efficient methods of screening and averting the dissemination of false narratives. The attempts to keep the information intact in the modern world of computer technologies could be enhanced with the assistance of the enhanced methodologies including social relationship aspects [9].

Neural network models suggested by Ma et al. (2015) and Wang et al. (2019) were proposed with the aim of tracing the rumors as they change with time. Ma et al. have designed a model, the Dynamic Series Time Structure model, that have taken into consideration social context features thereby making the model to outperform the previous models. Conversely, the neural solution introduced in the work by Wang et al. divided the propagation framework into different steps, depending on the time when tweets were posted, using content characteristics with an adaptive propagation model to determine a consistent system of rumor detection.[10].Such advances demonstrate the relevance of time in misinformation propagation understanding. The ability to emphasize the interrelationship between societal contexts and the timing; in which information is distributed is helping researchers to make the rumor detecting systems more accurate and reliable. These developments are essential in coming up with efficient mechanisms which can adequately reverse the flow of fake information in real-time in social media settings. [11].

Scholars Researchers have consulted the development of news distribution across the years in terms of the temporal factors and user relations. Do et al. (2019) presented the Dual RNN for Rumor Detection (DRRD) that relied on the features of users to predict the possible rumors. In the meantime, Liu et al. (2017) devised a model, which was based on the different user representations to differentiate rumors and authentic messages. This type of model was actualized on the Chinese Sina Weibo social media. Such undertakings emphasize the importance of knowing when and how rumors are spread socially. Scholars are progressing the limits of the rumor detection system by incorporating some of the user functions and various forms of representation. This kind of model is crucial when it comes to the enhancement of misinformation identification particularly in dynamic online spheres where fake stories can travel very fast [12]. The Labeled Propagation Tree (LPT) presented by Cao et al. (2017) targets automatic rumor detection. This model combines the propagation structure dynamics and user influence in addition to featuring novel elements. Through these new aspects of the detection process, LPT offers a better model in the formation of the rumor propagation processes within the context of the social networks. Besides enhancing the accuracy of rumor identification, this strategy also helps to draw attention to the importance of user interactions that determine the information flow. In view of the serious problems that misinformation presents to digital communication, these models will be of importance to designing impactful counter-measures and creating more trustworthy information ecosystem [13].

Researchers found that the likelihood to share doubtful breaking news is higher in less educated people and old ones (Glenski et al., 2017). Also, the audience that is attracted by clickbait, or the information related to conspiracy theories is at higher risk of spreading propaganda. These results demonstrate the necessity to study the demographics and user interests regarding the misinformation distribution. With the help of determining particular groups that are more vulnerable in sharing false information, researchers and policymakers will be able to work out specific interventions to propagate media literacy and critical thinking. Increasing awareness about vulnerable populations might contribute to minimizing the circulation of misleading information, which will eventually create a more informed society [14]. (2015) classified news into five steps, such as trusted, disinformation, propaganda, clickbait, and conspiracy theories. The activity of propagating propaganda and clickbait is to manipulate and redirect traffic such that the user can be duped and this fact alone must be highlighted as being critical abut ascertaining how the least important and most irrelevant of fake news is spread using the social media [15].

Transferring the information begins when one person posts a content, and then his proximate friends also post it. A research framework created by Wen et al. (2015) examines the behavior of users regarding both positive and negative information, focusing on the role of social factors in the process of spread dissemination. In research, emphasis has been laid on the influence of social dynamics in shaping the information flow throughout networks. The model provides insightful information on how information propagation through users and the kind of the content exchanged between them occurs. Being aware of these processes, strategies may be formulated to eliminate the misinformation and contribute to the overall quality of information spreading in social media settings. [16]. Natural Language Processing Semantic analysis is the aspect of Natural Language Processing using context to make a valuation of subjective and objective aspects in text. It aims to understand the meaning and ideas expressed in written text and categorise it into various categories, which include fraud, hate speech, satire, crime and misinformation. As it has been reported by Alnoso et al. (2021), there has been a growth in the popularity of fake news that is sometimes produced using a wide range of techniques associated with literature. Of paramount interest here is the introduction of effective tools to interface semantic analysis so as to detect and counter misinformation made possible by this evolution. Through proper sorting of material, they have a potential to moderate the effects of misleading stories, by giving the general population a better sense of clarity. With the current dynamism of information, there is a need to deepen our content on semantic contexts, in the process of creating media literacy and proper sharing of information. [17].

Air journal of Artificial Intelligence
ISSN: 3106-7743 | 3106-7735
Volume 3, Issue 1, 2026

However, Butani et al. (2019) approach the problem of fake news circulation by providing semantic analysis as a tool of revealing assorted forms of misleading news. They include the features of sentiment in their depiction of features. In spite of the fact that machine learning and natural language processing have been already used in the identification of the false news, the issue that seems to be investigated further should be the exploration of the other technologies that could be in use and the other fields that it could be applied to. In their work, more focus is on extraction and development of semantic understanding that represents an important aspect in the identification of fake news. Such emphasis on semantic interpretation is critical, since it may additionally enable the effectiveness of detection systems to be increased and enable them to distinguish between authentic and fraudulent materials. With the increasing spread of misinformation, it will be necessary to widen the research on different technologies and approaches to developing their robust solutions since they will be useful in effectively combating fake news propagation in a different setting [18].

The expansion of social media services has led to the wide spread of the political news causing the need in the automated fact-checking procedures. Such websites as Politifact, Snopes, Twitter and Chrome have developed mechanisms to check the accuracy of the information and fight misinformation. Machine learning, deep learning, and natural language processing techniques are used to identify false news, so the field is very useful to the readers who need credible news.

The need of proper fact-checking system proves more visible, as the landscape of information evolves. Such technologies do not only aid detecting lies but also allow users to deal with the peculiarities of information overload. The further developments in these directions will prove essential in improving the transparency of news and in developing a well-informed discourse, which will become part of a more healthy media environment [19].

Problem Statement:

The issue of the detection of fake news is of utmost importance because of its diverse nature. Most of the prior studies are centered on developing false news classification models without providing the effort to identify the rationale behind the same. Our primary concern in this principal issue is the determination of how we gear up the news content with maximum explicitness and identification of semantics features of the text that indicate novelty or deceitful characteristic in the content of the news. There is a demand to identify the single, reused, believable, and practical ontology that can identify fake news by relying on the hypothesis of semantic analysis with the support of machine learning classifiers. The problem of fake news identification is grave due to a variety of its features, such as rapid distribution of false information and the impact on the opinion of people. Our knowledge has a gap in the logic behind the mechanics of deception since the other research has been more geared towards creation of categorization models that do not provide us with any views on the reasoning behind false news. Main problems raised here by us are how to organize the news content and how it is possible to identify the textual semantic features that would point to uniqueness or dishonesty in the news. This is fundamental to investigate the wording and context connection of how a real news differs with the fake news. In order to detect false news, we should create a reusable, reliable and effective ontological approach combining the strengths of the machine learning classifiers and semantic analysis hypothesis.

Objectives:

Fake News Identification and Classification Using Machine Learning has as its main goal to come up with an effective and strong system in the identification and classification of fake news materials being produced by others as a comparison to the real information. The necessity to reduce the level of misinformation and disinformation, which might influence social, political, and economic aspects seriously, guides the given overall aim.

These sub-objectives are:

Semantic Analysis on the Context of News: Semantic analysis on the context of news is among the main goals of the given study. Semantic analysis is used to interpret meaning of words and sentences and as a result the system will not simply detect keywords. This approach to revealing certain signs of misinformation can be viewed as useful due to the possibility of detecting them by analyzing the context

in which a piece of news can be conveyed and thus, some of the signs cannot be manifested by more conventional methods. It will enable the system to pick on the linguistics of words like tone, intent (the intention behind the words) and implied meaning, which would help in creating a more realistic conclusion of whether a piece of news is credible or not. Due to this more in depth appreciation of the material then this will aid in helping to classify or differentiate between actual news as opposed to fake news, which in many cases is created to mislead or obscure by either distorting context or supplying information that is out of proportion. By conducting semantic analysis, the study will enhance fake news detection into greater precision and reliability.

Identification and Analysis of Main Entities and their Relationship in News Dataset: The identification of main entities and their relationships within a news dataset and finding their relationships is one of the most important tasks of fake news designing and classification with the help of machine learning. Entities are important elements of the news, i.e. people, organizations, places and happenings. In pointing out such entities, the system will be able to obtain information on the fundamentals of the news story. In addition, the study of the relationship of these parties to one another can be used to identify patterns, contradictions, or dubious connections that can be evidence of fake news. Such entity-relationship analysis will enable the machine learning model to have a better understanding of the structure and context of the news contents and have a more accurate result in identifying authentic and fake information. This finally increases accuracy and reliability of fake news detection.

Computing Dynamic Results by Providing Ontology-Based Approach: The keyword in the title of the research on fake news identification, namely, computing dynamic results by providing an ontology-based approach, is to understand that the study will make use of the structured, hierarchical knowledge in order to repackage an even better comprehension and classification of the news content. Ontology-based approach will enabled the system to portray relations between entities, concepts and attributes in a formal manner. With this approach, the system will be able to dynamically examine news stories and be able to read in to the inner correlation and sense of the textual material. Besides enhancing the level of accuracy in locating fake news, the methodology reduces the rigidity in adopting new changes in false information distributions and thus the operation becomes scenario-wise. This eventually results in more advanced and dependable findings since the ontology gives a basis of a more profound semantic perception and the identification of trends.

Semantic Features to Increase Accuracy: The main aim of the study will be to increase the accuracy of machine learning model to identify and classify fake news through the incorporation of semantic features. Semantic features do not break down to the surface-level characteristics of the news content, but they include analysis of the meaning and the context in which the news elements lie. The study concentrates on features of links between words, phrases, and entities in a piece of news in order to embrace a spectral pattern which may reflect the authenticity of a news or a fake one. Such a strategy will enable machine learning models to grow a deeper comprehension of the latent connections in the content and thus will enhance their capacity to recognize the truth quality in information. The inclusion of these semantic knowledge in the models makes the study more reliable and accurate when carrying out the fake news detection process.

Research Question and Hypothesis:

The above-mentioned research problems can lead us to some fake news problem-related questions that need to be solved by giving some effective performance-based fake news Identification model.

- RQ1. How to extract semantic features within news content to classify it as fake news?
- RQ2. How to compute relationships between entities by analysing the semantic knowledge?
- RQ3. How to provide an ontological approach with a machine learning classifier to infer news knowledge for making dynamic decisions?
- RQ4. How to classify fake news based on semantic analysis by using machine learning classifiers?

Literature Review:

In this section, we have mentioned some other authors' incredible work that has been done so far to detect, predict or identify fake information on social media and the impact of false news on society.

Ke wu, et al. (2015) proposed an approach based on graph kernel-based SVM [20]. Four types of features

Air journal of Artificial Intelligence
ISSN: 3106-7743 | 3106-7735
Volume 3, Issue 1, 2026

are extracted to detect false rumor Identification. Linguistic features (contains message text information length, punctuation, hashtag, etc.), User features (contains information about user gender, age, location, number of msgs posts, etc.), The graph-kernel-based classifier is developed, by using all the features it is used to find the similarity of the propagation tree. Graph theory is used to detect rumors in social networks. Users are denoted by nodes and their relationships are denoted by edges. Some of the nodes are considered as 'monitor nodes. Propagation of message patterns is used to find the best results. The proposed algorithm provides 90% accuracy of results. [21]

Zhifan Yang, et al. (2015) suggest a model to identify fake news in social networking websites by using a novel hot topic Identification method combined with a sentence modeling method. The proposed method includes the hot topic Identification with some important features like features based on content, based on Twitter and on the Network. Some binary classifier methods are used to detect rumors like Naive Bayes, Random Forest and Logistic Regression to test different features to detect whether the document is a rumor or not. [22]

Shihan Wang, et al. (2015) present a method to detect rumors based on patterns of rumors. Patterns are based on the structural and behavioral features of rumors. A graph pattern-based algorithm is proposed to match the patterns of features from social networking websites. This algorithm worked on both valid and false news and then distinguished between them. Rumor patterns differentiate between rumor and valid news. [23]

Yang Liu, et al. (2016) proposed a model based on diversified user presentation to find out the difference between rumors and valid messages. After working on message generation patterns their hypothesis idea proves to be right to differentiate between rumors from valid messages. SVM classifier is used to detect rumors. When any social media news proliferates among different literacy levels of users the proposed method combines the users in different groups according to their characteristics. The proposed model is for the Chinese social media site Sina Weibo [24].

Weiling Chan, et al. (2016) consider rumors as anomalies and propose an approach Factor Analysis of mixed Data to detect anomalies in social media. A Euclidean formula to calculate the distance and similarity of cosine is applied to find the degree of Deviation. Based on the degree of Deviation, ranks are provided to rumors and according to their rank rumors are detected. The proposed algorithm works well in Online social media. [25]

Ontology-based Fake News Identification:

Ontology is a branch of philosophy that describes and provides knowledge in the form of concepts within a specific domain and gives an understanding of the connection between entities. The purpose of ontologies is to model the terms within a domain and find the relationship between them and that can be applied in different areas of interest like knowledge management systems and the e-commerce domain. The benefits of using ontologies include these are reusable within the domain knowledge. The information used within the domain is understandable and reusable by different agents and people.

Our main area of research is Twitter (a social networking site where people interact and gain information about their interests) where we provide semantic feature extraction on Twitter news content and build an ontology to retrieve tweets and provide sentiment analysis on these tweets.

Chelsea et al. propose an approach by using the ontologies and related concepts that can help find accurate information for problems related to fake news Identification quickly and efficiently. The model consists of extracted information from tweets combined with a weighing formula. The model works in two steps (i) How to create ontology and (ii) How to weight the ontology to get appropriate information. The work model includes finding the relevant information that has been lost in the large datasets. The relevant retrieval information is then taken to use in the model to avoid damage in the field. The limitation of work includes what ontology to be used in the relevant field. The idea has been generated from the information taken from dissatisfaction with customs related to deliveries. A series of technologies from extracting tweets from Twitter, cleaning the data, the analysis subject, construction of an ontology, and sentiment analysis of tweets have been used. They have used Protégé software to build an ontology. The resulting analysis

ISSN: 3106-7743 | 3106-7735
Volume 3, Issue 1, 2026

is used by companies to take corrective measures to avoid dissatisfaction among users [26].

Rashkin et al. (2020) define the problem of the proliferation of fake news in the online environment as well as in political affairs. Linguistic features of news have been used for fact-checking of news [27]

Kontopoulos et al. (2014) propose the deployment of original ontology-based techniques towards a more efficient sentiment analysis of Twitter posts. The novelty of their work is not only characterizing the semantic score of the post but also achieving a semantic grade of distinct notion for a tweet post by applying machine learning classifiers. This results overall in a more elaborate analysis of post opinions regarding a specific topic. The purpose of this research is to analyze the sentiment within the post opinion. The area of research is related to the social networking site Twitter. The model of ontology is designed dynamically instead of explicitly defining the hierarchies of concepts within ontology. The ontology has been built by utilizing the existing designed ontologies to reduce the classes and redundancies in the model [28].

Comparative Analysis:

The comparative analysis below has been designed to show various factors within the research area of domain fake news Identification such as precision, accuracy, efficiency, reasoning, reliability, and recall. Precision means the model has a close degree of correctness when compared to the perfect model. Accuracy is as the model provides exact results. Efficiency refers to the model that provides the best results with minimum resources. Reasoning means providing the actual logic behind using any specific methodology. Reliability means the ability of the model to perform consistently in a correct manner. Recall can be defined as the value obtained by dividing the correct results by the number of results that need to be returned. Automatic machine refers to a machine not requiring human support for its operation.

Critical Analysis:

Most of the previous work has been done by extracting linguistic, structural, and temporal features of message propagation. Content-based approaches focus on text extraction from news and user profiles for rumor Identification. Some of the researchers used users' comments along with the content of the news to find rumors. Few used traditional rumor Identification approaches that are based on extracting features like users' context and propagation structure of the rumors etc. Besides some authors have provided better fake news Identification techniques but still require some additional information need to be extracted from the features of the context of the news. It has become considerably important to integrate other technologies into fake news Identification. Our research focuses on extracting and constructing the semantic understanding of the domain of fake news Identification. Very little work has been done in this scenario. After analyzing the comparative analysis of existing literature, we came to know that semantic analysis has not been done to analyze the news content and extract the meanings of text within the news. According to our knowledge before applying classifiers to the dataset we need to do a semantic analysis on the text to extract the conceptual meaning of the news context. Different authors have used different techniques to improve performance, efficiency, precision, accuracy, reasoning recall, etc.

Research Methodology:

Proposed Semantic Fake News Identification System:

As in the previous section, we have mentioned fake news Identification methodologies. According to the analysis of previous approaches the need of a semantic analysis-based approach that can classify fake news in a dynamic environment is required. Our goal is to provide a reliable, efficient, and reusable Fake News Identification System that can analyze text to extract semantic notation of news content to classify news. Considering that, we have proposed a semantic fake news Identification approach with ML Classifiers to answer the research questions stated in the previous section of the paper and to provide a conceptual understanding of content A Semantic Fake News Identification System has been divided into three phases. The first phase includes data acquisition in which we have gathered the data set according to the types of classifications of fake news. The second phase includes the semantic layer in which context features have been extracted from news content.

Air journal of Artificial Intelligence
ISSN: 3106-7743 | 3106-7735
Volume 3, Issue 1, 2026

Data Acquisition:
First of all, we took a dataset of fake news from kaggle.com. The dataset contains text and is gathered from 24 websites by using API webhose.io. The dataset designed specifically to analyze news to categorize it into fake or real. Fake news can be categorized into various concepts. Fake concept has further expanded into the following categories hate, fraud, conspiracy, crime, clickbait, satire, bias, etc. We have taken these categories as classes to classify fake news.

Ontology Builder:
The ontology builder is responsible for the semantic analysis of text data and the results generation. Ontology is like an organized data container to categorize data according to the domain knowledge and provide results as meaningful data and relationships of data instances by analyzing the semantics of data. The process is instantiated with ontology and through querying the model results are obtained. Ontology is built by using Protégé. Results obtained by ontology ha discussed in Chapter 4.

FNDOnt Description:
By applying semantic analysis to the fake news dataset, we have extracted features from news content. It provides us with the named entities and the relationship between these entities that have been extracted according to the type of classes. The extracted relationship has been mapped onto ontology. We have built ontology in the first step and a database connection has been established. To find reasoning from existing entities, we have applied the SPARQL query to the database. Protégé is used for ontology development in Ontology Web Language (OWL). For experimentation and simulation, a Fake news data set is used that has been acquired from the open-source data repository kaggle.com. We also developed the intermediate software component for OWL/XML to implement the fake news Identification system defined in the ontology. Moreover, we define inference-based semantic rules to analyze the text of news content while query generation is being done.

Classifier:
In this step, we try to draw some conclusions according to the classes of fake news into categories of fake or real. The machine Learning algorithm has been trained on the training dataset and the performance of the ML model has been checked by testing in the test dataset. We have used Random Forest, Logistic Regression, and LSTM in our model. We have used binary classification. Dataset has been classified into Fake or real news.

Random Forest:
ML model, Random Forest has been used for classification problems and regression problems as well. The classifier was introduced by Brieman. The records have shown in n number of trees. To utilize the subset of features different decision trees has created and each tree provides a separate output result. All trees work parallel without correlation of attributes or features which is why all trees are independent of each other. The trees have generated n number of outputs, which map into a single class to predict the final results based on average and probability. Using more trees means using more features provides higher accuracy to the model.

Logistic Regression:
The classifier is used to classify the dataset unless its name is regression but it also solves the classification problems. The output of the model gives the results in an "S" shape. The classifier used the sigmoid function to map the output values within probabilities. The classifier can find the variable that can be used to provide the best classification results. Logistic Regression helps in making decisions after understanding the relationships between features and makes predictions about the outcome of models. The outcome of Logistic Regression bounds between 0 and 1

LSTM (Long Short-Term Memory):
LSTM (Long Short-Term Memory) is capable of learning long-term dependencies between news articles. LSTM can memorize information over a long period. The workflow of LSTM consists of cell states and a connections between cell states. The flow of information from cells uses multiple gates. Gates used a sigmoid function with a multiplication operator. LSTM isn't able to change the information

ISSN: 3106-7743 | 3106-7735
Volume 3, Issue 1, 2026

between the flows of information. So the first step of LSTM is to decide which information to keep and which information to discard. The actual information that will pass through LSTM is the first step of the process. The decision will occur by the forget gate, also called the sigmoid layer. The output of the forget gate is either 0 or 1. If the output is 0, information is discarded, and 1 means keeping the information completely.

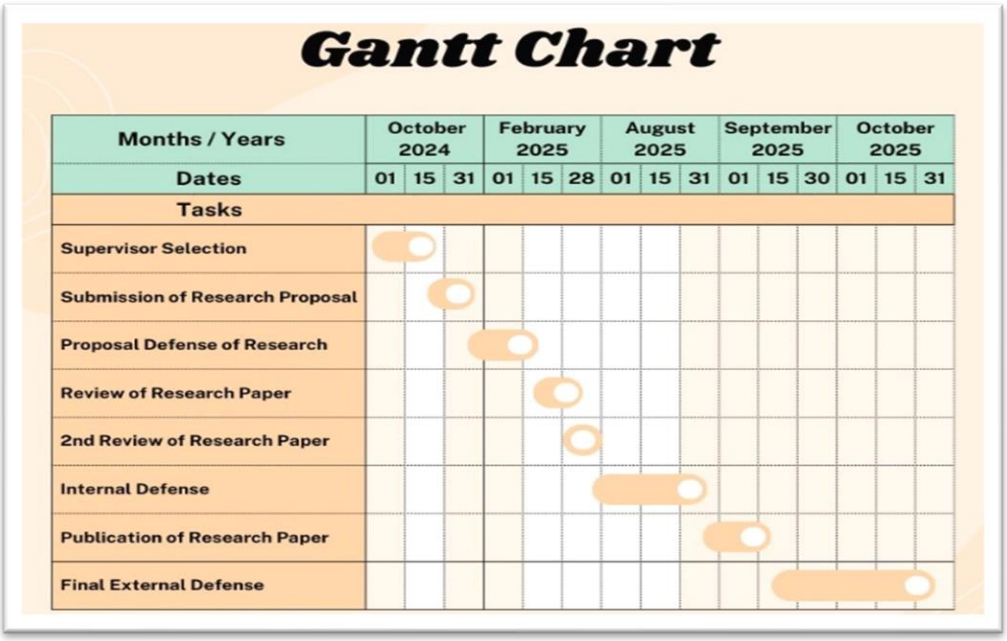
Proposed SFNDS using ML:

In this paper, we have proposed a novel approach based on semantic feature extraction from news articles and use this extracted knowledge to classify fake news. The main purpose of our research is to use semantic features in Machine Learning models to improve the accuracy of models. The model is shown in the extracted entities. Relationship between entities mapped to find the semantic understanding of text semantic analysis on the news content of the fake news dataset has applied. The extracted labeled dataset that has gained from FNDOnt is the input of our ML model. Data has been prepared to import in ML classifiers. In natural language processing, feature selection from dataset to model on our classifiers is an important step. The more important features have been selected the more the performance of the model would be. Text is converted into vector space that has some features.

Limitations and Scope:

Although machine learning has the potential to detect and categorize the fake news, the effectiveness of the expected solutions has some limitations. The business problem regarding the quality and the diversity of the training data is a major one. The problem is that fake news datasets tend to be one-dimensional (they do not vary in terms of language, cultural background, and a variety of misinformation) and that this leads to biased or incomplete models. Moreover, fake news is dynamic and evolves to change the trends of misinformation and induce artificial intelligence models so that the rapid pattern replacement within fake news is not possible. The other restriction is that the texts are solely used and although the semantic analysis brings good results, it can leave unnoticed the factor of imagery, videos, or other multimedia content that contributes immensely to the spread of fake information. Also, the computational requirements of the integration of ontology-based techniques and semantics may become a problem in the face of real-time processing thus restricting the system to handle large volumes of data or high-velocity social network like twitting. This research is not confined to ordinary fake news detection but moves on to enhance the level of credibility of all information on social media because of use of advanced machine learning methods. With a combination of semantic analysis and entity relationship modeling, it will be possible to explore the content of news in more detail and find deeper interconnections and contradictions that will not be noticeable when using conventional text-based forms of reasoning.

Air journal of Artificial Intelligence
ISSN: 3106-7743 | 3106-7735
Volume 3, Issue 1, 2026



Gantt Chart:

REFERENCE

[1] S. Zamani, M. Asadpour, and D. Moazzami, "Rumor identification for Persian tweets," in Proc. 25th Iranian Conf. Electrical Engineering (ICEE), 2017.

[2] Y. Liu and Y.-F. B. Wu, "Early identification of fake news on social media through propagation path classification with recurrent and convolutional networks," in Proc. 32nd AAAI Conf. Artificial Intelligence (AAAI-18), 2018.

[3] D. Mrowca and E. Wang, "Stance identification for fake news identification," unpublished.

[4] G. Giasemidis, N. Kaplis, I. Agrafiotis, and J. R. C. Nurse, "A semi-supervised approach to message stance classification," IEEE Trans. Dependable and Secure Computing, 2018.

[5] M. R. Islam, S. Muthiah, and N. Ramakrishnan, "RumorSleuth: Joint identification of rumor veracity and user stance," in Proc. IEEE/ACM Int. Conf. Advances in Social Networks Analysis and Mining, 2019.

[6] S. Wang, Q. Kong, Y. Wang, and L. Wang, "Enhancing rumor identification in social media using dynamic propagation structures," in Proc. IEEE Conf., 2019.

[7] J. Du, R. Xu, Y. He, and L. Gui, "Stance classification with target-specific neural attention networks," in Proc. 26th Int. Joint Conf. Artificial Intelligence (IJCAI-17), 2017.

[8] W. Chen, C. K. Yeo, C. T. Lau, and B. S. Lee, "Behavior deviation: An anomaly identification view of rumor preemption," in Proc. IEEE Conf., 2016.

[9] B. Malhotra and D. K. Vishwakarma, "Classification of propagation path and tweets for rumor identification using graphical convolutional networks and transformer-based encodings," in Proc. 2020 IEEE Sixth Int. Conf. Multimedia Big Data (BigMM), 2020.

[10] J. Du and R. Xu, "Stance classification with target-specific neural attention networks," in Proc. 26th Int. Joint Conf. Artificial Intelligence (IJCAI-17), 2017.

[11] Y. Wang, S. Qian, J. Hu, Q. Fang, and C. Xu, "Fake news identification via knowledge-driven multimodal graph convolutional networks," in Proc. ICMR '20, Dublin, Ireland, 2020.

[12] K. Wu, S. Yang, and K. Q. Zhu, "False rumors identification on Sina Weibo by propagation structures," in Proc. IEEE Conf., 2015.

[13] Y. Liu and Y.-F. B. Wu, "Early identification of fake news on social media through

ISSN: 3106-7743 | 3106-7735
Volume 3, Issue 1, 2026

propagation path classification," in Proc. AAAI Conf., 2018.
[14] Z. Yang, C. Wang, F. Zhang, Y. Zhang, and H. Zhang, "Emerging rumor identification for social media with hot topic identification," in Proc. IEEE Conf., 2015.
[15] S. Volkova, K. Shaffer, J. Y. Jang, and N. Hodas, "Separating facts from fiction: Linguistic models to classify suspicious and trusted news posts on Twitter," in Proc. Int. Conf., Vancouver, 2017.
[16] S. Wang and T. Terano, "Detecting rumor patterns in streaming social media," in Proc. 2015 IEEE Int. Conf. Big Data, 2015.
[17] S. Wang, I. Moise, D. Helbing, and T. Terano, "Early signals of trending rumor event in streaming social media," in Proc. 2017 IEEE 41st Annual Computer Software and Applications Conf., 2017.
[18] Y. Liu and S. Xu, "Detecting rumors through modeling information propagation networks in a social media environment," IEEE Trans. Computational Social Systems, 2015.
[19] S. Jain, V. Sharma, and R. Kaushal, "Towards automated real-time identification of misinformation on Twitter," in Proc. 2016 Int. Conf. Advances in Computing, Communications and Informatics (ICACCI), Jaipur, India, 2016.
[20] A. Dey, R. Z. Rafi, S. H. Parash, S. K. Arko, and A. Dey, "Fake news pattern recognition using linguistic analysis," in Proc. 2018 Joint 7th Int. Conf. ICIEV & 2nd Int. Conf. icIVPR, 2018.
[21] Y. Qin, D. Wurzer, and C. Tang, "Predicting future rumours," Chinese Journal of Electronics, vol. 27, no. 3, 2018.
[22] C. M. M. Kotteti, X. Dong, and L. Qian, "Multiple time-series data analysis for rumor identification on social media," in Proc. 2018 IEEE Int. Conf. Big Data, 2018.
[23] M. Vohra and M. Kakkar, "Identification of rumor in social media," in Proc. IEEE Conf., 2018.
[24] Z. Huang, H. Yao, and Y. Wu, "Rumors identification in Sina Weibo based on text and user characteristics," in Proc. 2018 IEEE Advanced Information Management, Electronic and Automation Control Conf. (IMCEC), 2018.
[25] M. Aldwairi and A. Alwahedi, "Detecting fake news in social media networks," in Proc. 9th Int. Conf. Emerging Ubiquitous Systems and Pervasive Networks (EUSPN), 2018.
[26] M. Amith and C. Tao, "Representing vaccine misinformation using ontologies," Journal of Biomedical Semantics, vol. 9, no. 22, 2018.
[27] M. Bharti and H. Jindal, "Automatic rumour identification model on social media," in Proc. 2020 Sixth Int. Conf., 2020.
[28] E. Kontopoulos, C. Berberidis, T. Dergiades, and N. Bassiliades, "Ontology-based sentiment analysis of Twitter posts," Expert Systems with Applications, 2013.