OPEN ACCESS

Kashf Journal of Multidisciplinary Research
Vol: 01 - Issue 12 (2024)
P-ISSN: 3007-1992 E-ISSN: 3007-200X
https://kjmr.com.pk

A SEMANTIC FAKE NEWS DETECTION SYSTEM USING MACHINE LEARNING CLASSIFIER

Muhammad Rashad*
Department of Basic Sciences Superior University, Lahore

Noreen Khalid
School of CS & IT
Beaconhouse National University

Ameer Hamza
Superior University, Lahore (Pakistan)

Sikander Javed
Department of CS & IT Superior University, Lahore (Pakistan)

Kashif Bilal Majeed
Superior University, Lahore (Pakistan)

Corresponding Author: Muhammad Rashad, muhammad.rashad@superior.edu.pk

Article Info

Abstract

The purpose of fake news detection system is to build ontology to find hypothesis involved in misleading social media users through automated reasoning. Ontology for classification of news content has been created after understanding the semantic notations of textual features with in fake news dataset. The dataset we have used in our approach openly available on open-source data repository with the name fake News. The proposed model will provide semantic analysis of news content of the dataset and classification of news content into fake categories. The outcome of our proposed solution can be originating by applying three different classifiers of machine learning that is Random Forest, Logistic regression and LSTM (Long Short-Term Memory) that showed results about fake news and the accuracy of our proposed methodology is almost about 97%, 98% and 99% respectively. Thus the results prove that machine learning models performed better after analyzing the semantic features from news datasets.

Keywords: *Social Media, Fake News, Propagation of news, Semantic Analysis, Ontology, Machine Learning.*

Introduction

Social networking websites have become one of the biggest sources of information (Vosoughi et. al, 2018). Nowadays with the help of Social networking websites information can be delivered within no time and can be targeted to a large number of audiences (Park & Chang, 2017). With the intrusion of digital media in our personal life some malevolent factor is also involved. It is necessary to put some measures on the authenticity of the news and classify them into true and false news that we are getting from social media sites (Duffy et al., 2020). Semantic analysis of news content is getting more intention from researches to understand the meaning of the news context and by using semantic of text to classify news into fake or real category.

Fake news or misinformation can cause distress and anxiety in society (Aceto et al., 2019). With the increased usage of social media people tend to take news from digital media more than any other traditional sources of information (Visentin et al., 2019). This characteristic of social media triggers the problem of spreading any rumor to gain someone's personal benefit. Social networking websites such as Facebook, Twitter, Instagram and Sina Weibo lack tools that can easily detect rumors and to stop spreading them. It becomes more difficult now when usage of data in social media has enormously increased. Detection of rumors in social media is one of the most challenging parts in the data science field (Bondielli & Marcelloni, 2019). Deep learning algorithms have proven to be better where users have been interrelated with thousands of users all over the globe (Zhang et al., 2019). These approaches widely used in classification applications with the increased amount of data (Zhan et al., 2022). Working with images, texts and sounds was not possible before deep learning. The main advantage of this technique is that manual feature extraction is not required. The model has been trained in a format that works automatically. With the rising speed of fake news some automated techniques need to take over the artificial and machine learning approaches.

Detection of fake news is a critical problem due to its various aspects (Zhang & Ghorbani, 2020) Previous researches mainly focus on the building classification models without giving intention to inferring the reasoning involved behind false news (Zhang & Ghorbani, 2020). The main problem we identify here is how to gear up news content in an explicit manner and to identify the semantic characteristics of text that shows originality or deception in the news content. We need to find a reusable, credible and efficient ontological approach that can detect fake news by using the hypothesis of semantic analysis with the help of machine learning classifiers.

The main objective of this study is to find out fake news on social networking site Twitter by analyzing the semantic features of news content.

- Improve the credibility of information provided by social media sites.
- Providing semantic analysis on context of news
- Finding main entities and their relationship from news dataset.
- Computing dynamic results by providing ontology based approach.
- Providing an approach to leverage semantic features to enhance the accuracy of machine learning models.

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 detection model.

1. How to extract semantic features with in news content to classify it into fake news?
2. How to compute relationships between entities by analyzing the semantic knowledge?
3. How to provide an ontological approach with machine learning classifier to infer news knowledge for making dynamic decisions?
4. How to classify fake news on the basis of semantic analysis with machine learning classifiers?

Our contribution regarding fake news detection work includes building an ontology that's functionality can enhance with the passage of time with ML classifier. Ontology has been structured to include the rapidly proliferating news contents on twitter and to classifying the fake news from real news. Our proposed approach consists of three layers. In the first step, a dataset related to fake news has been collected. We take news content information that has been spread on twitter. Preprocessing of text data has been done to remove special characters and removing errors from dataset. Preprocessing has been done manually by using Modern CSV file editor application. Semantic features have been extracted from news contents and in the second step named entities and their relationships have been mapped. Ontology model that has built for the domain of fake news detection can acquire the knowledge constituted in it, and by using the ontological model we can attain our desired results within domain of fake news detection. Our proposed approach has been focused on to figure out dynamically classification for fake news.

This paper is organized into five chapters. The 2nd chapter consits of literature review on the basis of earlier research workings and chapter 3 describes data collection and methodology. In fourth chapter, the results and discussion are presented. Finally, chapter 5 presents conclusion and recommendations for future work.

1. Related Work

The research article focuses on integrating semantic analysis with machine learning classifiers to enhance fake news detection. This review synthesizes relevant studies that inform the methods presented in the paper, focusing on hybrid deep learning approaches, rumor detection, and semantic analysis.

Nasir et al. (2021) introduced a hybrid CNN-RNN model to combine the spatial feature extraction of CNN with the sequential processing of RNN, improving fake news detection by considering both content and temporal features. Moreover, Kumar et al. (2021) surveyed various machine learning and deep learning techniques, emphasizing NLP methods for text-based detection, which are key to the semantic detection system proposed in this research. In case of rumors detection mechnism, Wu et al. (2020) developed a propagation graph neural network (GNN) with an attention mechanism to detect rumors based on their spread in social media networks. To identify false information, Bharti and Jindal (2020) proposed a model for automatic rumor detection using machine learning, which also considers the structure of social media conversations. For the alignment with the semantic analysis approach of the proposed system, Konkobo et al. (2020) suggested a deep learning model for early detection of fake news on social media, using CNNs for feature extraction. Also, Umer et al. (2020) introduced a hybrid CNN-LSTM model for stance detection, combining CNN for feature extraction and LSTM for sequence learning, which enhances the detection of fake news through deep learning techniques.

Malhotra and Vishwakarma (2020) utilized graphical convolutional networks (GCNs) to classify rumor propagation paths, emphasizing the importance of analyzing how rumors spread in networks. In 2020, Mandical et al. highlighted the role of machine learning classifiers in identifying fake news, which is similar to the proposed system's use of classifiers for text-based detection. Tiwari et al. (2020) discussed the use of machine learning algorithms such as decision trees, support vector machines, and random forests for detecting fake news based on text features, highlighting the role of machine learning in accurate detection. For fake news detection, Wang et al. (2019) proposed a multimodal graph convolutional network (GCN) combining textual and visual data, and emphasized the importance of incorporating external knowledge to improve detection accuracy. Kesarwani et al. (2019) examined the use of the k-nearest neighbor (KNN) classifier for fake news detection on social media, demonstrating how similarity-based classification can be effective in distinguishing fake content. Groza (2020) presented an ontology-based approach for detecting COVID-19-related fake news, utilizing semantic reasoning and domain-specific knowledge to enhance detection accuracy. Sharma and Sharma

(2019) discuss the challenges of fake news detection, emphasizing the need for advanced techniques to address misinformation's evolving nature. Their work supports the proposed system's aim to improve accuracy through semantic analysis. Henry and Stattner (2019) focused on predictive models for early hoax detection on Twitter, highlighting the importance of early identification. Their approach aligns with the proposed system's goal to detect fake news early using semantic analysis. Lahlou et al. (2019) reviewed automatic fake news detection methods, underscoring the need for hybrid techniques that combine linguistic analysis, social network analysis, and machine learning. This aligns with the proposed system's integration of semantic and machine learning techniques. Srinivasan (2019) introduced a parallel neural network approach for faster rumor detection in social networks. This approach is relevant to the proposed system's use of neural networks for efficient detection. Islam et al. (2019) introduced *Rumorsleuth*, a model that detects both rumor veracity and user stance. This could complement the proposed system by analyzing both content and user context for better fake news detection. Yang et al. (2019) presented an unsupervised approach that detects fake news without labeled data, which could help the proposed system adapt to new types of fake news. Wang et al. (2019) enhanced rumor detection by considering dynamic propagation structures in social media. This approach aligns with the proposed system's aim to track how information spreads and evolves across platforms. Castelo et al. (2019) developed a topic-agnostic method to identify fake news, providing a more flexible detection system. The proposed system could integrate similar techniques to detect fake news across various topics. Reshi and Ali (2019) reviewed methods for detecting rumor proliferation on social media. Their work underscores the need for effective systems to track and identify fake news, which aligns with the proposed system's focus on rumor detection and semantic analysis. Giri and Sachdeva (2019) explored anomaly detection in social networks, highlighting techniques that can identify irregular patterns indicative of fake news. This complements the proposed system's use of machine learning to detect anomalous content in social media. Monti et al. (2019) applied geometric deep learning models to detect fake news on social media. Their graph-based approach aligns with the proposed system's focus on understanding news propagation patterns and detecting fake news using graph-based learning. Amith and Tao (2018) discussed the use of ontologies to represent and classify vaccine misinformation. This methodology can inform the semantic aspect of the proposed system, particularly in domain-specific fake news detection, such as health misinformation. Kotteti et al. (2018) explored time-series data analysis for rumor detection on social media, emphasizing the importance of tracking temporal patterns for identifying fake news progression over time. Mahid et al. (2018) provided a review of various detection techniques on social media, highlighting the use of machine learning classifiers for distinguishing fake news, supporting the proposed system's approach. Dey et al. (2018) focused on linguistic analysis to recognize fake news patterns, aligning with the proposed system's use of natural language processing to detect fake news based on linguistic features. Munir and Anjum (2018) discussed the use of ontologies for effective knowledge modeling and information retrieval, which can enhance the semantic understanding of content, aiding in more accurate fake news detection. Qin et al. (2018) focused on predicting future rumors by analyzing patterns in social media data, which aids in understanding the spread of fake news. This method helps in recognizing the progression of rumors and in early detection. Liu and Wu (2018) proposed using recurrent and convolutional networks to classify fake news based on its propagation path. Their approach emphasizes the early detection of fake news by tracking how it spreads across platforms. Aldwairi and Alwahedi (2018) applied machine learning techniques to detect fake news in social networks. This work supports the use of classifiers in the proposed semantic detection system, aiming to differentiate between real and fake content based on network features. Ma et al. developed a kernel learning model to identify rumors on microblogs by analyzing propagation structures. This methodology enhances the proposed system by incorporating structural analysis to detect and categorize fake news spread. Thota et al. explored deep learning approaches for fake news detection, utilizing feature learning techniques to classify news content. This aligns with the use of deep learning in the proposed system for more accurate and efficient detection of fake news.

Table 1: Comparison of Related Work Technique

Paper Title	**Reliability**	**Precision**	**Recall**
An Ontology-Supported Misinformation Model: Toward a Digital Misinformation Library	✓		
A Topic-Agnostic Approach for Identifying Fake News		✓	
Ontology-Based Sentiment Analysis of Twitter Posts	✓		✓
Representing vaccine misinformation using ontologies	✓		
Sentiment Analysis tweets	✓	✓	✓
Ontology-based Sentiment Analysis Process for Social Media	✓		✓
The Use of Ontologies for Effective Knowledge Modeling and Information Retrieval	✓		✓

3. Proposed Methodology

In this section, we will make sentiment analysis of the procedure to detect fake news.

3.1 Proposed Semantic Fake News Detection System

A Semantic News Detection System using Machine Learning Classifier leverages semantic analysis to extract contextual features from news content, enabling deeper understanding and classification of misinformation (Tiwari et al.,2020). By combining semantic knowledge with machine learning algorithms, this system ensure reliable, efficient, and adaptable detection of fake news in dynamic digital environments (Mandical et al., 2020).

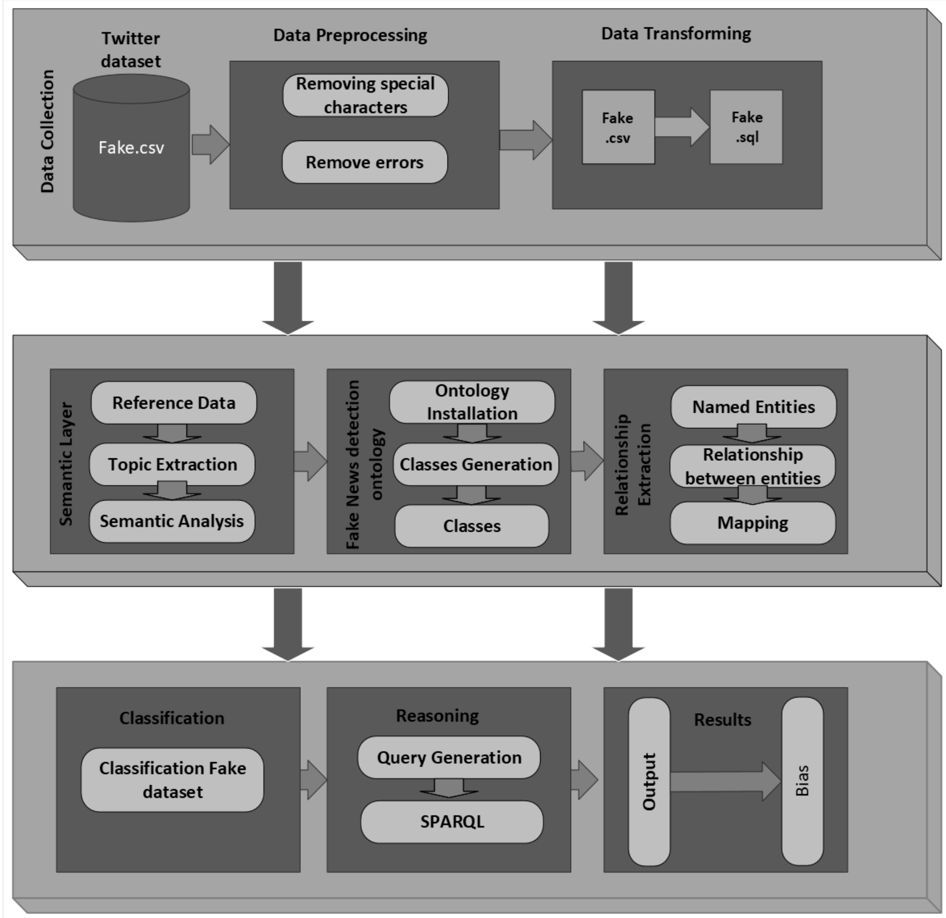

Figure 1: Flowchart of proposed approach

3.2(a) Data Acquisition

First of all, we took a dataset of fake news from kaggle.com. The dataset contains text and gathered from 24 websites by using API webhose.io. Dataset designed specifically to analyze news to categorize them into fake or real. Fake news can be categorized into various concepts. Fake concept has further expanded into following categories hate, fraud, conspiracy, crime, clickbait, satire, bias etc. We have taken these categories as classes to classify the fake news. The extracted data properties contain information about fake news, author of the article, text and title of the news has been used to extract sentiment of news. There are two type of reviews about news has extracted such as users' share count and likes count. Comment count is a numeric property. News content can be of positive, negative or neutral, emotion or attitude about an aspect of the entity from an opinion holder.

(b) Ontology builder

The proposed system uses ontology based on Object-Oriented Programming (OOP) principles to model relationships in news content. Objects are represented as nouns, and their attributes or actions as verbs, with categories defined for accurate classification. Designed using Protégé, the ontology is tailored for fake news detection in the dynamic social media environment, where misinformation spreads rapidly. By categorizing data into meaningful clusters, the system performs semantic analysis to identify and understand patterns in news content. The integration of machine learning classifiers enhances detection accuracy, making the system a reliable and scalable solution for combating fake news.

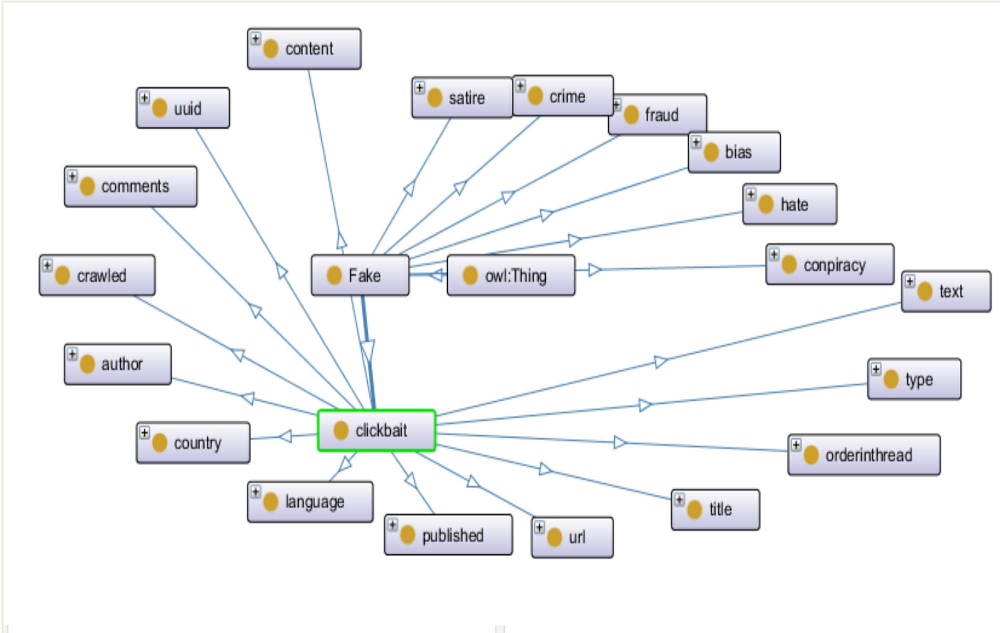

Figure 2: Fake News Detection Ontology

The ontology builder is responsible for semantic analysis on text data and the results generation. Ontology is like an organized data container to categorize data according to the domain knowledge and providing results as meaningful data and relationships of data instances by analyzing semantics of data.

(i) FNDOnt description

The proposed Semantic Fake News Detection System uses semantic analysis to extract relevant features from a fake news dataset, focusing on named entities and their relationships based on predefined classes. These relationships are mapped onto an ontology, which is built using Protégé and implemented in Ontology Web Language (OWL). SPARQL queries are applied for reasoning from existing entities stored in a database. The system follows a three-step process: data collection, semantic analysis, and classification.

In the first step, the dataset is acquired from Kaggle, cleaned, and processed using Modern CSV to make it compatible with MySQL through an API. The second step involves establishing a connection via JDBC and uploading the dataset to Protégé for ontology development and evaluation. In the final step, SPARQL queries are used for reasoning and categorizing news content, including the identification of fake news categories.

The ontology includes four key classes: Social Media Platform, Social Media User, News Content, and Classification. The system integrates sentiment analysis to detect fake news, with the relationships between entities manually defined to ensure accurate classification. This approach ensures a reliable, dynamic, and scalable solution for fake news detection.

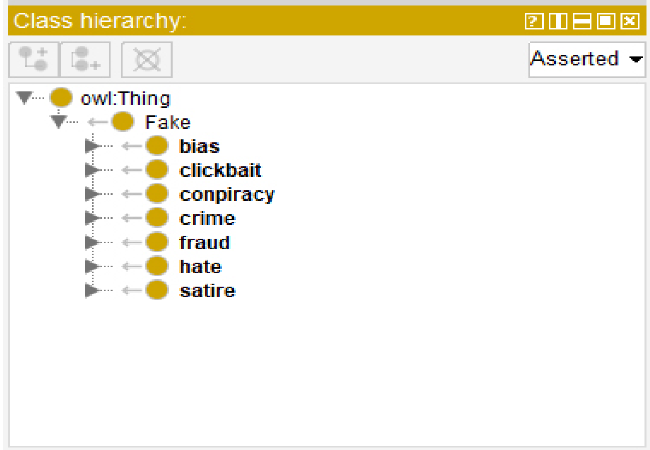

Figure 3: Fake news detection class hierarchy

In the system, object properties represent relationships between instances or individuals. For fake news detection, classes such as "users," "news content," "social media platform," and "classification" are defined. The object property "has relation" links users to other users, while "connect" represents the relationship between users and social media platforms. Additionally, the object property "subclass of" is used to define hierarchical relationships between these classes.

(c) Classifier

In this section, we will first introduce the machine learning classifiers that we have used in our model.

Random Forest

ML model Random Forest has used for classification problems and regression problems as well. The records have shown in *n* number of trees. To utilize the subset of features different decision trees has created and each tree provide a separate output results. All trees works parallel without correlation of attributes or features that is why all trees are totally independent with each other. The trees have generated *n* number of outputs and these outputs map into single class to predict the final results on the basis of 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 model gives the results in "S" shape. The classifier used sigmoid function to map the output values within probabilities. The classifier has the ability to find the variable that can be used to provide best classification results. Logistic Regression helps in making decisions after understanding the relationships between features and makes the 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 is able to memorize information from a long period of time. The working flow of LSTM consists of cell states and a connection between cell states. The flow of information from cells uses multiple gates. Gates used a sigmoid function with a multiplication operator. LSTM aren't able to change the information between the flows of information. So the first step of LSTM is to decide which information to keep and information to discard. The actual information that will pass through LSTM is the first step of the process. Decision will take place by the forget gate also called sigmoid layer. The output of the forget gate is either 0 or 1. If the output is 0 it means information is discarded and 1 means to keep the information completely.

Proposed SFNDS using ML

This paper introduces a novel approach for fake news detection through semantic feature extraction from news articles, aimed at improving machine learning model accuracy. The primary goal of the research is to incorporate semantic features into machine learning models, enhancing their performance. The extracted entities and relationships are mapped to provide a semantic understanding of the text, which is then analyzed using semantic techniques. The labeled dataset obtained from the FNDOnt ontology serves as the input for our machine learning model.

In this study, feature selection plays a crucial role in training classifiers, with the most significant features chosen to optimize model performance. We employ TF-IDF (Term Frequency-Inverse Document Frequency) for text vectorization, which converts textual data into numerical form while preserving the word sequence. TF-IDF assesses term frequency within documents and measures their importance, with higher scores indicating more relevant terms.

The dataset, processed using the scikit-learn library, is split into training (70%) and testing (30%) sets. Three machine learning classifiers Random Forest, Logistic Regression, and LSTM are used to compare the results. Our experiments demonstrate that the Random Forest classifier outperforms the others in terms of accuracy, making it the most effective model for semantic-based fake news detection.

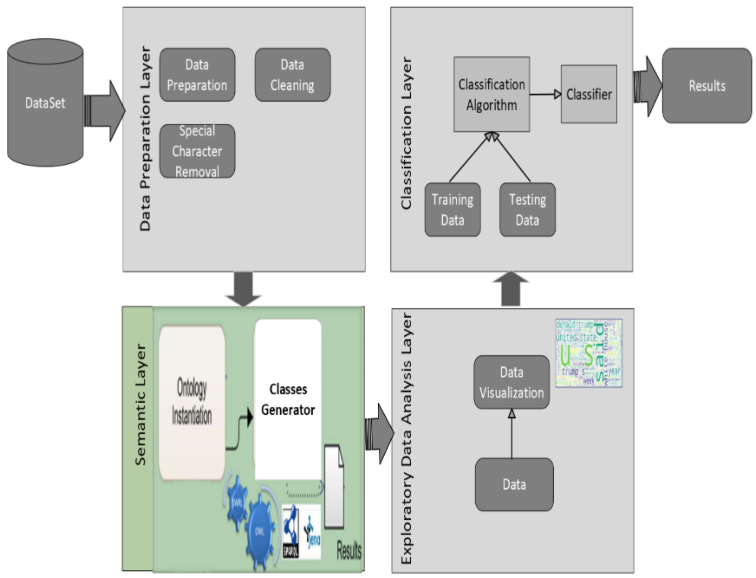

Figure 4: Proposed Semantic Fake News Detection System using ML

4. Results and Disscussions

In this chapter, we make the semantic rules for dynamic fake news detection. We make the SPARQL queries for the validation of results. We configure Reasoner from the reasoner menu and start reasoning for valid results.

Hate Result

uuid	author	type	published	title	text
017f86b2045577d795cdfd671a8157ee2d7c0f77	Ian	hate	2016-10-28T01:13:01.130+03:00		If this is really true our leaders are sic
088ddf580942133f16728aa4773e94bcc26bcfc0	apoc36	hate	2016-10-27T02:54:41.397+03:00	Kids being made to vote for President in School	Kids being made to vote for President
098f5a30ec280282544ed77a837cb2d1d873d9f0	bluuder	hate	2016-10-28T01:04:00.000+03:00		He is alive: https://www.youtube.con
0f06e348dee64d1e9aecdb381937c99e3b445b53	Kim Petersen	hate	2016-10-30T02:00:00.000+02:00	An Eight Point Brief for Voting to Avoid Corpora...	22 Shares 1 20 0 1 The anarchist polit
1017fec72846254fbb3441529f5f912ec902029c	Jason Ditz	hate	2016-10-29T06:02:35.504+03:00	Amnesty Warns Against Using White Phosphoru...	Notes Evidence of Use Northeast of M
11df9bd055c59241dbb7a38b9b90a3091154cc47	ahtribune.com	hate	2016-10-27T03:00:00.000+03:00	Executes 190 Iraqi Civilians in Hammam al-Alil, 4...	ISIS executes over 230 civilians in Mo
13b2a3aa1853d50f8f4353986fa702ce50ab0a0d	Robsi Micheal	hate	2016-10-29T09:34:00.000+03:00		Brilliant!
13e3e1983787c76876e9e5d095e73ef8cee1628a	Scout Bishop	hate	2016-11-10T08:41:00.000+02:00	TOMI LAHREN Has Special Message For Celebrit...	Copyright © 2016 100PercentFedUp.
16272301eef853acec79c45bd273e916ff4b9caf	hqanon	hate	2016-10-28T03:00:00.000+03:00	Political Party of Anarchists, Libertarians, Hacke...	By Claire Bernish at thefreethoughtpr
162b39e6518af4e591acdd65d719b4f858596b3e	EdJenner	hate	2016-11-12T19:55:23.883+02:00	#BoycottComedian..ROBERT DENIRO Wanted ...	Go to Article Donald Trump was willing
16f19ffbbfe178bbf72a2a6699a02c147b330a46	righteous	hate	2016-10-28T03:00:00.000+03:00	What is Anonymous?	What Anonymous is and what it mean
17c06cb24ac5d32042fab31d8cb9b8b80d31404c	ahtribune.com	hate	2016-11-02T02:00:00.000+02:00	Iranians Spend $2.1b on Beauty Products Annu...	11 Shares 1 9 0 1 Iranians spend 4.5
1952f74ad6fd40a47501b5e055b3c596ee0147be	hqanon	hate	2016-10-27T03:00:00.000+03:00	'Dillary Trumplon': This Guy's Epic Rap Song Just...	By Nick Bernabe at theantimedia.org
1b9104850bec39ed6a90d185913108c87f267767	Ivan Jose	hate	2016-11-06T20:03:00.000+02:00	SNL Gets Real And Delivers The Most Important...	on November 6, 2016 8:03 am · Satu
1c32f441f24a42ed4072b07bb314c078937bcb47	Robert	hate	2016-10-27T08:54:00.000+03:00		Fraud fraud fraud can't anyone see t
1dc6ff5a17dbaae65f902edc49ea475a4ceb892a	Rebecca Ben...	hate	2016-11-22T07:58:00.000+02:00	Trump ERUPTS At Secret Meeting With The Pres...	on November 21, 2016 7:58 pm · Do
1e956fb8691f0534fc21f70e769b5d9b44e983e7	willz	hate	2016-10-27T02:54:59.244+03:00	Nanobots causing overwhelming depression. No...	Nanobots causing overwhelming depr
1fd7c9a718708972572b7246e6f445e0ad835049	Ivan Jose	hate	2016-11-04T23:25:00.000+02:00	Former Classmate: Trump Smacked His Son So ...	on November 4, 2016 12:25 pm · We
20e1a45f760f8db5f26207bbfe827bf8cd6d65bb	Ivan Jose	hate	2016-10-27T09:18:46.193+03:00	Comment on Tutorial: Riding The Philippine Jeep...	adobochron 1 Comment MANILA, Phil

Figure 5: Hate News Output

- **Introduction**

uuid	author	type	published	title	text
0206b54719c7e241ffe0ad4315b808290dbe6c0f	Mason Hugh	bias	2016-11-01T21:56:00.000+02:00	FANTASTIC! TRUMP'S 7 POINT PLAN To Reform...	Email HEALTHCARE REFORM TO MAK
063726c998dfdbe9e917b618f205b7aa7bbffcef	ActivistPost	bias	2016-11-03T19:17:10.178+02:00	Some Want Injustice To Be Distributed Evenly, ...	By Vin Armani The federal landgrab p
096858c81b7e65acb92a0070dc7c37ac70c49a44	Activist Post	bias	2016-11-03T20:33:06.135+02:00	Smart Meter Case Testimony Before the Pennsy...	By Catherine J. Frompovich This is t
09a6989efa0ae4b702bf7e35b00a9902ec959485	Activist Post	bias	2016-11-04T18:00:50.740+02:00	Iraqi Car Bomb Jaw-Dropping Crisis Actor Faker...	By Bernie Suarez A recent video rele
10b7f0b943835bccb69d5ca90c4ee2c1e2570954	Activist Post	bias	2016-11-07T18:25:56.671+02:00	FBI Wants you to Believe It Examined 650,000 ...	By Claire Bernish In no surprise to an
160453f5b7386ad17e4c7c7882f15b1346dfb497	Mason Hugh	bias	2016-10-28T20:39:00.000+03:00		I dont know guys , i must say that at
17359768953b61269780878e26d074a37f679b7c	Brandon Turbeville	bias	2016-11-05T16:35:21.682+02:00	Drone Restrictions Can Help Peaceful Protester...	By Shane Trejo The Dakota Access P
175c7828a1a86c7ff20bf993539e82e6ed5c2fdd	Brandon Turbeville	bias	2016-11-04T19:56:01.728+02:00	U.S. Elections "November Chaos": What You're ...	By GRTV The FBI's October surprise
27d1cf9b1c84474e17b0166fb5985151e8e95a31	Brandon Turbeville	bias	2016-11-08T17:20:22.420+02:00	9 Reasons Why I Am NOT Voting	By Chris Duane Chris Duane explains
2bdc29d12605ef9cf3f09f9875040a7113be5d5b	reasoning with facts	bias	2016-10-29T08:47:11.259+03:00	Re: Why Did Attorney General Loretta Lynch Pl...	Why Did Attorney General Loretta Ly
2ca77a341a3e26e3f842f2804aef1ecf62ba90f6	Mason Hugh	bias	2016-11-05T18:01:00.000+02:00	California Secessionists to Meet at Capitol Day ...	By Joseph Jankowski An organizatio
2ee23dfd1bc388c251fec378231fe3bbd8ef0883	Anonymous	bias	2016-10-29T10:38:00.000+03:00		Good Day AD I would have to say th
2f836e2554e5b1b7533ea25b3461b06e38c5355b	Brandon Turbeville	bias	2016-11-08T18:00:23.026+02:00	Cops Fire Tear Gas on Water Protectors as The...	By Justin Gardner As the controversi
32b8a10c927682a608bd3b88147c357b8807eaeb	tokyowashi (norepl...	bias	2016-10-29T15:58:39.811+03:00	FBI Weiner Probe Reopens Hillary Clinton Inves...	EXCLUSIVE #Breaking FBI Reopens I
3a7ac4cdd4a98abe359e36b2c62650cc6bfda11c	Brandon Turbeville	bias	2016-11-09T01:16:24.080+02:00	The U.S./Turkey Plan For "Seizing, Holding, And...	By Brandon Turbeville As the U.S. Pr
4729f48f5301dd29fa27cfc59c3c50f8b298d848	Activist Post	bias	2016-11-05T18:11:59.518+02:00	False Flag? US Intelligence Warns Of Likely Al Q...	By Whitney Webb Over the course c
47d021d9d6d842cafaba334e0930c5b49f9ea154	Activist Post	bias	2016-11-08T22:56:55.077+02:00	What Is "Fake" News?	By Catherine J. Frompovich One of t
48ed7ee6e668ecfecfbe59144d8b4192d9bd7474	Brandon Turbeville	bias	2016-11-07T23:05:53.804+02:00	Police Caught Spying on Journalists to Uncover ...	By Chris "Kikila" Perrin In what can or
4d3faf17519cfa46ca754ab80684288 18dbb2bf3	EdJenner	bias	2016-11-07T22:34:36.822+02:00	YIKES! HILLARY GOES OFF THE RAILS...Pulls A ...	
4f34dd134c678cd33a0906d509b407e4cf6ef9d9	Activist Post	bias	2016-11-05T17:30:36.699+02:00	Smart Meter Case Testimony Before the Pennsy...	By Catherine J. Frompovich This is th
525633ff8631f5b245a6d09141722d7c264c5a5b	John Wick	bias	2016-11-07T18:06:13.864+02:00	Here's a Summary Of the Engineered US Electio...	By Bernie Suarez It's always importar

Figure 6: Bias News Output

Crime Result

uuid	author	type	published	title	text
00ecd7fc7097c0078eabe3f6dcb3077da971f47e	Henry Wolff	crime	2016-11-05T00:25:00.000+02:00	Soros Spends $2 Milion to Defeat Arpaio	Soros Spends $2 Million to Def
00fb73e408d13e789d4b74787a079b715477a261	sebastien Provencher	crime	2016-10-27T07:59:52.571+03:00		I'm ready come to me , you wi
02ae76c658c0a52244fddeaf9c14b5d0dde4e9d7	Activist Post	crime	2016-10-30T02:35:25.843+02:00	Like a "Concentration Camp" Police Mark DAPL P...	By Claire Bernish On Thursday
03702848a0914899c4f5e8e7836659b10757a5b4	Jason Ditz	crime	2016-11-04T04:16:14.946+02:00	Iraq PM Threatens War Over Turkey's Military B...	Says Iraq Will Deal With Turke
03905a0f06cb711190aad3934af253a0e1c00a83	newyorkgiant	crime	2016-10-26T23:12:00.000+03:00		So ,you have Rothschild bank
046dc3d06f7181f4e9e7e0ba7d0bfc521befe988	Alex Ansary	crime	2016-10-31T19:54:37.595+02:00	China repeatedly hacked US, stole data on nuk...	China repeatedly hacked US,
0548a4a7a5d414eab993531b26f053fd92e6b76f	Activist Post	crime	2016-11-01T04:46:02.654+02:00	China's Airport Security Robot Can Deliver Elect...	By Nicholas West While debat
06908df0e86b9ae22a52feac79776c641d8ec586	Jason Ditz	crime	2016-11-04T04:16:09.463+02:00	Assange Confirms: WikiLeaks Didn't Get Emails F...	Accuses Clinton Campaign of
07517581586f16e842364402f43ed60f3a472858	Brandon Turbeville	crime	2016-10-28T07:35:29.388+03:00	U.S. Takes A Stab At A No Fly Zone In Two Plac...	By Brandon Turbeville As the
0a2e9889ef324c5c323498b2045227488a885538	Ryanmorph	crime	2016-10-27T00:44:00.000+03:00		N379P / Piper PA-46-350P Mal
0c4ece3343fc14de9d5c7ae5c0bd4f3e1bf228e5	Henry Wolff	crime	2016-11-04T06:01:00.000+02:00	Family Remembers Queens Sucker Punch Victim ...	Family Remembers Queens Su
0c9d01b2bc7e6e5be2ebecd52b54468b5bf1e731	Alex Ansary	crime	2016-11-02T19:24:00.901+02:00	Warning or threat? Hillary hints at second civil w...	Warning or threat? Hillary hint
0d16b7b8853e18fde62e354b863a0f0d05f95cfe	Alex Ansary	crime	2016-11-04T02:44:42.831+02:00	Doug Casey: A Civil War Could Be in the Cards ...	Doug Casey: A Civil War Coul
0d599ada25a336c0df13ee689cd612154a5e9a51	Activist Post	crime	2016-10-29T04:20:38.885+03:00	Distracted By Election 2016, No One Resisted T...	By Nathaniel Mauka Congress
0f79cbbd7220380b8ec2511152fee88ca5e568ce	Margaret Griffis	crime	2016-11-03T09:23:02.737+02:00	Sectarian War Crimes Reported near Mosul; 28...	Share This Islamic State leade
10165357c27399c0b85c74c0830238239b3902c3	hqanon	crime	2016-10-26T03:00:00.000+03:00	Putin blasts Clinton & Tells U.S. Govt to STOP C...	In the clip – filmed in Russia –
11ecd71be0955967a9058366b7bee2499dbc3e95	Robsi Micheal	crime	2016-10-26T22:19:00.000+03:00		Trump has an excuse now to
1272ae8f77556a0b90d0ca088db19ee7ff2d13f5	burdman30ott6	crime	2016-10-28T06:57:23.266+03:00	To anyone relying on the not enough troops to i...	To anyone relying on the "not
144fc0dba9ee289bf18660cf906c1e08cfaeb749	Alex Ansary	crime	2016-11-07T20:01:08.682+02:00	Donald Trump kicks off final campaign day with ...	Donald Trump kicks off final ca

Figure 7: Crime News Result

Confusion Matrix of Machine Learning Classifiers

Confusion matrix used to show the results of machine learning models. The output has shown in four parts TN, FN, FN and TP.

The output matrix of Random Forest has shown below.

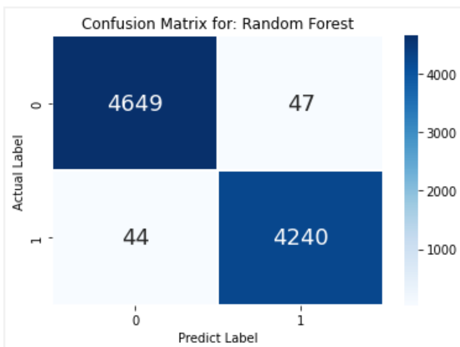

Figure 8: RF Confussion Matrics.

Output of Logistic Regression has shown in figure below.

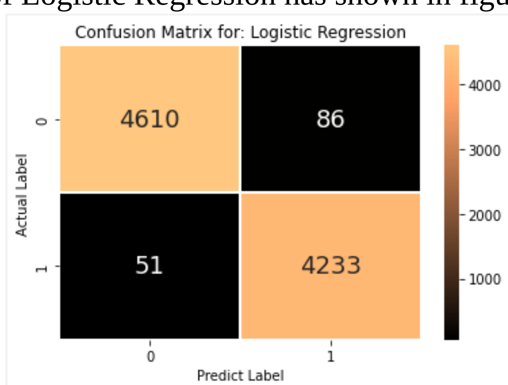

Figure 9: LR Confussion Matrics.

The output of LSTM model has shown in figure below.

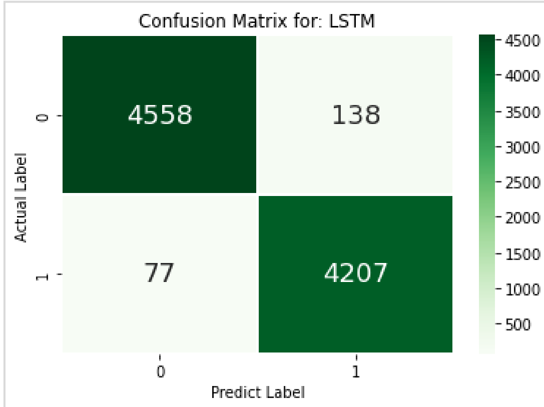

Figure 10: LSTM Confussion Matrics.

ROC curve

To show the output results in graphical form ROC curve has used. The ROC curve contains two parameters True Positive Rate and False Positive Rate.

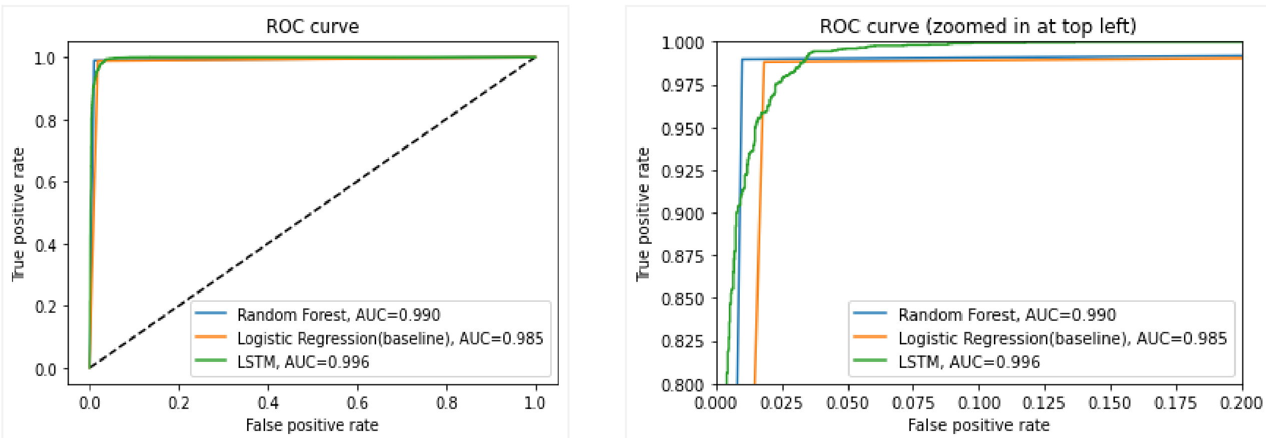

Figure 11: ROC Curve Positive.

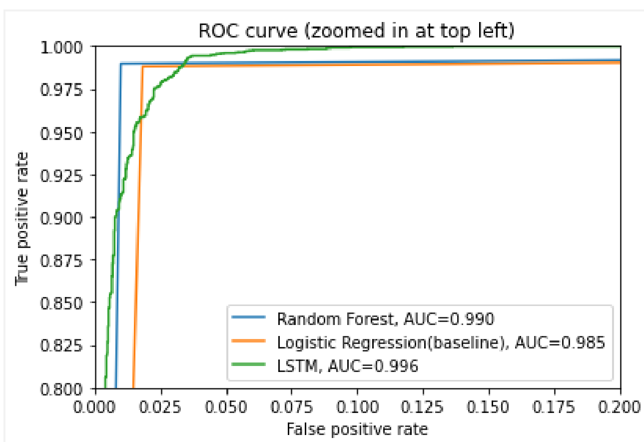

Figure 12: ROC Curve Negative.

Our research enhances Fake News Detection by integrating semantic features into ML classifiers. The results show that semantic analysis improves model accuracy. We used an ontological approach to extract semantic features, which were then combined with Random Forest, Logistic Regression, and LSTM models. Each model was trained for 10 epochs. Random Forest achieved the highest accuracy in our proposed methodology.

3. Conclusion and future work

Our research presents an optimal solution for detecting fake news on social media, addressing the spread of harmful content. We propose an ontology-based system that uses semantic analysis to improve news classification. The model follows three steps: first, semantic features are extracted from news content and mapped to an ontology to establish relationships between classes; second, semantic rules are applied to infer knowledge from the context for dynamic fake news detection; and third, the ontological model is tested with machine learning classifiers such as Random Forest, Logistic Regression, and LSTM, with Random Forest achieving 99% accuracy. Future work aims to explore Deep Learning classifiers to further evaluate the performance of the proposed ontological approach.

References

Aceto, G., Ciuonzo, D., Montieri, A. & Pescapé, A. ''Mobile encrypted traffic classification using deep learning: Experimental evaluation, lessons learned, and challenges,'' IEEE Trans. Netw. Service Manag., Vol. 16, No. 2, pp. 445–458, Feb. 2019.

Aldwairi, M. & Alwahedi, A. "Detecting fake news in social media networks," Procedia Computer Science, Vol. 141, pp. 215-222, Jan 1, 2018

Amith, M. & Tao, C. "Representing vaccine misinformation using ontologies," Journal of biomedical semantics, Vol. 9, pp. 1-3, Dec, 2018.

Bharti, M. & Jindal, H. "Automatic rumour detection model on social media. Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)," IEEE, pp. 367-371, Nov 6, 2020.

Bondielli A, Marcelloni F. "A survey on fake news and rumour detection techniques," Inf Sci, 497, pp. 38–55, 2019.

Castelo, S., Almeida, T., Elghafari, A., Santos, A., Pham, K., Nakamura, E. & Freire, J. "A topic-agnostic approach for identifying fake news pages," InCompanion proceedings of the 2019 World Wide Web conference, pp. 975-980, May 13, 2019.

Dey, A., Rafi, R.Z., Parash, S.H., Arko, S.K. & Chakrabarty, A. "Fake news pattern recognition using linguistic analysis," In2018 Joint 7th International Conference on Informatics, Electronics & Vision (ICIEV) and 2018 2nd International Conference on Imaging, Vision & Pattern Recognition (icIVPR), pp. 305-309, Jun 25, 2018.

Duffy, A., Tandoc, E., & Ling, R. "Too good to be true, too good not to share: The social utility of fake news," Information, Communication & Society, *23*.13, pp.1965–1979, 2020.

Giri, V.K. & Sachdeva, S. "Anomaly Detection in Social Networks," In2019 9th International Conference on Cloud Computing, Data Science & Engineering (Confluence), pp. 698-703, Jan 10, 2019.

Groza, A. "Detecting fake news for the new coronavirus by reasoning on the Covid-19 ontology," arXiv preprint arXiv:2004, Apr 26, 2020.

Henry, D. & Stattner, E. "Predictive models for early detection of hoax spread in twitter. In2019 International Conference on Data Mining Workshops (ICDMW)," IEEE, pp. 61-64, Nov 8, 2019.

Islam, M.R., Muthiah, S. & Ramakrishnan, N. "Rumorsleuth: joint detection of rumor veracity and user stance," InProceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, pp. 131-136, Aug 27, 2019.

Kesarwani, A., Chauhan, S.S. & Nair, A.R. "Fake news detection on social media using k-nearest neighbor classifier. In2020 international conference on advances in computing and communication engineering (ICACCE)," IEEE, pp. 1-4, Jun 22, 2020.

Konkobo, P.M., Zhang, R., Huang, S., Minoungou, T.T., Ouedraogo, J.A. & Li, L. "A deep learning model for early detection of fake news on social media. 7th International Conference on Behavioural and Social Computing (BESC)," IEEE, pp. 1-6, Nov 5, 2020.

Kotteti, C.M., Dong, X. & Qian, L. "Multiple time-series data analysis for rumor detection on social media," In2018 IEEE international conference on big data (Big Data), pp. 4413-4419, Dec 10, 2018.

Kumar, S., Yadav, P. & Bagri, M. "A survey on analysis of fake news detection techniques. International conference on artificial intelligence and smart systems (ICAIS)," IEEE, pp. 894-899, Mar 25, 2021.

Lahlou, Y., El Fkihi, S. & Faizi, R. "Automatic detection of fake news on online platforms: A survey. In2019 1st international conference on smart systems and data science (ICSSD)" IEEE, pp. 1-4), Oct 3, 2019.

Liu, Y. & Wu, Y.F. "Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks," InProceedings of the AAAI conference on artificial intelligence, Vol. 32, No. 1, Apr 25, 2018.

Mahid, Z.I., Manickam, S. & Karuppayah, S. “Fake news on social media: Brief review on detection techniques,” In2018 Fourth international conference on advances in computing, communication & automation (ICACCA), pp. 1-5, Oct 26, 2018.
Malhotra, B. & Vishwakarma, D.K. “Classification of propagation path and tweets for rumor detection using graphical convolutional networks and transformer based encodings. Sixth international conference on multimedia big data (BigMM),” IEEE, pp. 183-190, Sep 24, 2020.
Ma, J., Gao, W. & Wong, K.F. “Detect rumors in microblog posts using propagation structure via kernel learning,” Association for Computational Linguistics.

Mandical, R.R., Mamatha, N., Shivakumar, N., Monica, R. & Krishna, A.N. “Identification of fake news using machine learning. International conference on electronics, computing and communication technologies (CONECCT),” IEEE, pp. 1-6, Jul 2, 2020.
Monti, F., Frasca, F., Eynard, D., Mannion, D. & Bronstein, M.M. “Fake news detection on social media using geometric deep learning,” arXiv preprint arXiv:1902, Feb 10, 2019.
Munir, K. & Anjum, M.S. “The use of ontologies for effective knowledge modelling and information retrieval,” Applied Computing and Informatics, Vol. 14, No. 2, pp.116-126, Jul 1, 2018.
Nasir, J.A., Khan, O.S. & Varlamis, I. “Fake news detection: A hybrid CNN-RNN based deep learning approach,” International Journal of Information Management Data Insights, Vol. 1, No. 1, Apr 1, 2021.
Park, Chang Sup, “Do social media facilitate political learning? Social media use for news, reasoning and political knowledge”, The Journal of Social Media in Society, Vol.6, No.2, pp.206-238, 2017.
Qin, Y., Dominik, W. & Tang, C. “Predicting future rumours,” Chinese Journal of Electronics, Vol. 27, No. 3,pp. 514-520, May, 2018.
Reshi, J.A. & Ali, R. “Rumor proliferation and detection in Social Media: A Review,” In2019 5th international conference on advanced computing & communication systems (ICACCS)’ pp. 1156-1160), Mar 15, 2019.
Sharma, S. & Sharma, D.K. “Fake News Detection: A long way to go. In2019 4th International Conference on Information Systems and Computer Networks (ISCON),” IEEE, pp. 816-821, Nov 21, 2019.
Srinivasan, S. “A parallel neural network approach for faster rumor identification in online social networks,” International Journal on Semantic Web and Information Systems (IJSWIS), Vol. 15, No. 4, pp. 69-89, Oct 1, 2019.
Tiwari, V,, Lennon, R.G. & Dowling, T. “Not everything you read is true! Fake news detection using machine learning algorithms. 31st Irish signals and systems conference (ISSC),” IEEE, pp. 1-4, Jun 11, 2020.
Umer, M., Imtiaz, Z., Ullah, S., Mehmood, A. & Choi, G.S. “Fake news stance detection using deep learning architecture (CNN-LSTM),” IEEE, Vol. 26, No. 8, pp.695-706, 2020
Visentin, M., Pizzi, G., & Pichierri, M. “Fake news, real problems for brands: The impact of content truthfulness and source credibility on consumers’ behavioral intentions toward the advertised brands,” Journal of Interactive Marketing, 45. 1, pp. 99–112, 2019.
Vosoughi, S., Roy, D., & Aral, S. “The spread of true and false news online,” *Science. 359*.6380, pp. 1146–1151, 2018.
Wang, Y., Qian, S., Hu, J., Fang, Q. & Xu, C. “Fake news detection via knowledge-driven multimodal graph convolutional networks. In Proceedings of the 2020 international conference on multimedia retrieval,” IEEE, pp. 540-547, Jun 8, 2020.
Wang, S., Kong, Q., Wang, Y. & Wang, L. “Enhancing rumor detection in social media using dynamic propagation structures,” In2019 IEEE International Conference on Intelligence and Security Informatics (ISI), pp. 41-46, Jul 1, 2019.
Wu, Z., Pi, D., Chen, J., Xie, M., & Cao, J. “Rumor detection based on propagation graph neural network with attention mechanism,” Expert systems with applications, Vol. 158, Nov 15, 2020.

Yang, S., Shu, K., Wang, S., Gu, R., Wu, F. & Liu, H. “Unsupervised fake news detection on social media: A generative approach,” InProceedings of the AAAI conference on artificial intelligence, Vol. 33, No. 01, pp. 5644-5651, Jul 17, 2019.
Zhan, Z., Li, J., & Zhang, J. “Evolutionary deep learning: A survey,” Neurocomputing, Vol. 483, pp. 42-58, 2022.
Zhang, C., Patras, P. & Haddadi, H. "Deep Learning in Mobile and Wireless Networking: A Survey," *IEEE Communications Surveys & Tutorials*, Vol. 21, No. 3, pp. 2224-2287, thirdquarter 2019.
Zhang, X., Ghorbani, A. A., “An overview of online fake news: Characterization, detection, and discussion,” Information Processing & Management, Vol. 57, No. 2, 2020.