Fake News Detection: A Supervised Learning Approach

Author

Tayyaba Rasool

FALL 2016-MS-16(CSE) 00000171719

MS-16 (CSE)

Supervisor

Dr. Wasi Haider Butt

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY

ISLAMABAD

January, 2019

Fake News Detection: A Supervised Learning Approach

Author

Tayyaba Rasool

FALL 2016-MS-16(CSE) 00000171719

A thesis submitted in partial fulfillment of the requirements for the degree of

MS Software Engineering

Thesis Supervisor:

Dr. Wasi Haider Butt

Thesis Supervisor's Signature: ____

DEPARTMENT OF COMPUTER & SOFTWARE ENGINEERING

COLLEGE OF ELECTRICAL & MECHANICAL ENGINEERING

NATIONAL UNIVERSITY OF SCIENCES AND TECHNOLOGY,

ISLAMABAD

January, 2019

DECLARATION

I certify that this research work titled *"Fake News Detection: A Supervised Learning Approach"* is my own work under the supervision of Dr. Wasi Haider Butt. This work has not been presented elsewhere for assessment. The material that has been used from other sources has been properly acknowledged / referred.

Signature of Student

Tayyaba Rasool

FALL 2016-MS-16(CSE) 00000171719

LANGUAGE CORRECTNESS CERTIFICATE

This thesis is free of typing, syntax, semantic, grammatical and spelling mistakes. Thesis is also according to the format given by the University for MS thesis work.

Signature of Student

Tayyaba Rasool

FALL 2016-MS-16(CSE) 00000171719

Signature of Supervisor

ACKNOWLEDGEMENTS

I would like to express my gratitude to my supervisors **Dr. Wasi Haider Butt and Dr. Arslan Shaukat** for their tremendous support, guidance and motivation.

I would also like to thank my Guidance and Evaluation Committee Members **Dr. Usman Akram** and **Dr. Urooj Fatima** for providing guidance and recommendations throughout the research phase. Their contribution has significantly improved my work.

Finally, I would like to express my gratitude to all the individuals who have rendered valuable assistance to my study.

Dedicated to my parents

ABSTRACT

The technological advances and the expansion in the digital scenario of the world in precious decade have brought about the capability of rapid information transfer over vast geographical regions. The rapid transfer of information on one hand has brought relief to human race in many aspects. At the same time the world as global village face many challenges too due to this rapid transfer of information. Among the challenges that have accompanied the revolution in information transfer is the exchange of misinformation. Rapid spreading of misinformation is a growing concern worldwide as it has the capacity to greatly influence not only individual reputation but societal behaviours. The consequences of unchecked spreading of misinformation can not only vary from political in nature to financial, but can also effect global opinion for a long time. The damage of the menace of fake news, being spread around in any form, can be beyond the imagination of human mind and its outcomes everlasting for the generations to come. Thus, detecting fake news is extremely important as well as challenging because the ability to accurately categorize certain information as true or fake is limited even in human. Moreover, fake news are a blend of correct news and false information making accurate classification even more confusing. In this paper, we propose a novel method of multilevel multiclass fake news detection based on relabelling of the dataset and learning iteratively. We tested our algorithm on metadata, text and a combination of both. The proposed method outperforms **the benchmark with an accuracy of 39.7% but maximum accuracy is achieved by holdout method using SVM classifier that is 66%.** Our experiments indicate that profile of the source of information contributes the most in fake news detection.

Table of Contents

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction

CHAPTER 1: INTRODUCTION

The introduction chapter gives comprehensive information of the thesis topic. It is divided into following portions. Section 1.1 provides the overview of fake news detection, Section 1.2 explains the problem statement being addressed in dissertation, Section 1.3 consists of research flow that is followed to complete the research, Section 1.4 is refers to the research contribution and Section 1.5 has the thesis organization.

1.1. Overview

In today's age, information transfer is very rapid, thanks to the internet and everything built around it. With faster information exchange, there is also faster misinformation exchange. Rapid spreading of misinformation is becoming a worldwide challenge because misinformation is the most effective weapon in the modern conflict. Online spread of misinformation has powers in shaping people's opinions and sentiments [1].

1.1.1. Introduction to Fake News

Fake news is a type of misinformation spread, via traditional media or social media, as factually accurate, to mislead audience[1]. Fake news on any media have the capability to harm a society or individual in different ways. It can have costly consequences in term of financial and political decision makings as well as destroy someone's social status. One of such incident happened after the Las Vegas shooting, when false information began to circulate and a picture of comedian Sam Hyde was shared on social media claiming he was the gunman[2]. In recent times, sharing fake news (or news without verification) has become a common practice.

Fake new is not a new problem, different groups use news media for propaganda and influence other groups. The faster information exchange makes fake news more powerful, especially since the US elections 2016. The open nature of internet makes it easier to create and spread fake news. Hence, fake news spread faster and deeper [2], due to which detecting fake new is very important but also technically very challenging.

Psychological foundation

Human beings are not good at identifying truthfulness of a news. There are several researches that can prove the psychological and cognitive power of fake news. Fake news

[1] https://en.wikipedia.org/wiki/Fake_news

[2] https://www.vanityfair.com/news/2018/01/the-6-fakest-fake-news-stories-of-2017

mainly aims to exploit the decision making of the audience. Due to some cognitive biases in human nature, fake news can often be perceived as accurate. There are mainly two reasons that make audience to believe in the fake news.

1. Audience tend to believe that their perspective is the accurate information, while others who disagree are considered as naïve. [3]
2. Audience prefer to receive the information that confirms their view. [4]

Psychological studies show that it is hard to correct the misinformation, when misperception is formed. Correction by presenting fact or accurate news is not helpful to reduce misperception but sometimes it increases the misperceptions, especially when there are groups of different perspective [4].

Social Foundation

Considering the entire news consumption ecosystem, we can also describe some of the social dynamics that contribute to the proliferation of fake news. Prospect theory describes decision making as a process by which people make choices based on the relative gains and losses as compared to their current state [5, 6]. This desire for maximizing the reward of a decision applies to social gains as well, for instance, continued acceptance by others in a user's immediate social network. As described by social identity theory[7, 8] and normative influence theory[9, 10] this preference for social acceptance and affirmation is essential to a person's identity and self-esteem, making users likely to choose socially safe options when consuming and disseminating news information, following the norms established in the community even if the news being shared is fake.

1.1.2. Fake News Detection

There are many tasks related to fake news detection, such as rumor detection [11] and spam detection [12]. Fake New detection aims to predict the probability of a news to be intentionally fake[13]. Detecting fake new is very important but technically very challenging. Firstly, the challenge is due to the fact that human beings are not so good at categorizing true and fake news. Human being, by rough comparison, can achieve only 50-63% success rate in identifying fake news [14]. Secondly, usually fake news are the blend of correct news with false information. It can easily confuse the audience and get their attention without noticing the fabricated information.

Fake News Detection Techniques: Categorization of detection algorithms can be determined by feature extraction of news. Fake news detection techniques mainly rely on news content and some extra social context information. Thus, it can be said that fake news are detected on the basis of news context and social context[15]. News context describes the information related to piece of news. Some news context attributes are headline, source, detail and visual content. On the other hand, social context feature can be derived from the social engagement of news. Social context features may be related to the audience, network of news propagation and reaction to the news.

1.1.3. Machine Learning

Machine Learning is the discipline in computer science that gives learning power to systems without having them comprehensively programmed [16-18]. The idea behind machine learning is to create intelligent algorithms that can be trained from any given set of data. On the basis of which, it can predict useful results. Machine leering emerged from pattern recognition and has some conceptual basis from artificial intelligence. It is also related to mathematics and statistics. Nowadays, it has wide application in variety of tasks that involve complex calculation and programming [19, 20]. It is very efficient in implementation of algorithms that are proved almost unworkable with good performance and quality. Our day-to-day-use applications such as voice recognition systems, social media facilities, video and audio surveillance, email filtering, finding online frauds etc. largely revolve around machine learning techniques [21-23]

1.2. Problem Statement

The challenge is due to the fact that human beings are not so good at categorizing true and fake news. Human beings, by rough comparison, can achieve only 50-63% success rate in identifying fake news [14]. Secondly, usually fake news are the blend of correct news with false information. It can easily confuse the audience and get their attention without noticing the fabricated information. Therefore, Multiclass problems are yet to be studied. The work of this dissertation is focused on solving the multiclass fake news detection problem using multilayer supervised learning.

1.3. Research Flow

The research process is carried out in a systematic way. Firstly we identify the problem, then we move to problem solving phase. A detailed study is carried out to for literature review that became the foundation of proposed solution. The proposed methodology provides a multilayer approach for the detection of fake news.

For this purpose, SVM and decision tree classifiers are used for training the models on the data set. After implementation, results are validated. Finally, the conclusion is drawn from results and future work is given.

1.4. Research Contribution and Main Objectives

This research is performed to develop a multi-label fake news detection process. A multilevel supervised learning approach is followed to improve the process. Following are the main objectives of this work:

- We develop a method that extracts features in order to classify news using multilevel supervised learning approach. The aim of the thesis is to provide a better algorithm that improves the accuracy of multi-label classification problem of news.
- We develop and train an input data set consisting of multi-labels and speakers profile (name, designation, party affiliation, credit history etc.) using machine learning classification techniques.
- We also develop and train and input dataset consisting of news statements using machine learning classification algorithms.
- We compare our approach with techniques and prove that our approach gives higher accuracy.

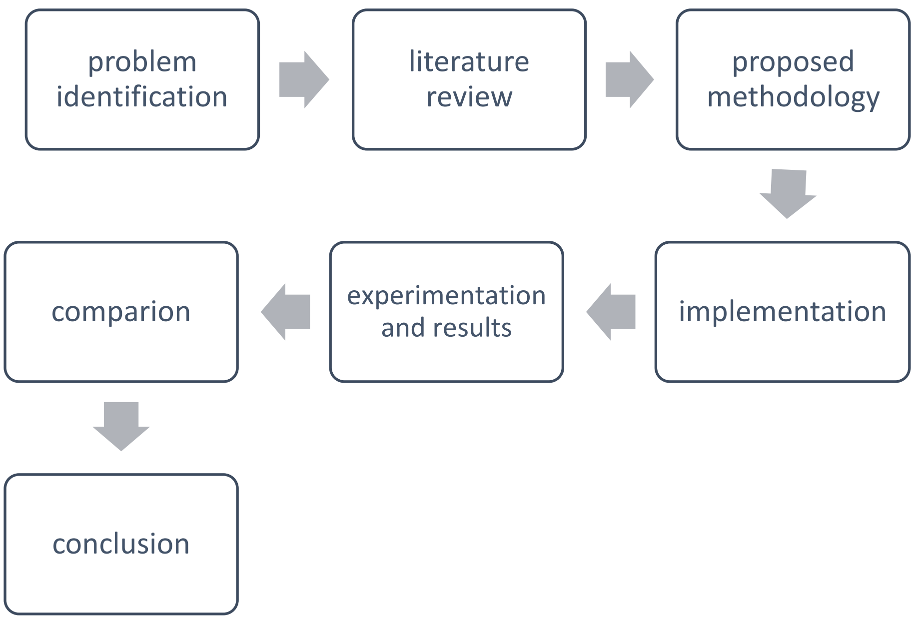

Figure 1.1: Research Design

1.5. Thesis Organization

Thesis organization is discussed using Figure 1.2. CHAPTER 1 goes through brief introduction enlisting the overview and background knowledge of topic, problem statement, objectives and contributions and then thesis organization. CHAPTER 2 has the detailed literature review. It delivers the work done in the field of fake news detection. Its first section contains many parts including the research methodology in which there is detail of category definition (in which five different categories are defined), acceptance and rejection criteria for the research study, the whole research process showing the step by step procedure for collecting the research data, quality assessment to verify the quality standard of review and finally the data extraction and synthesis. Second part comprehensively analyses the data on the basis of features and techniques used. The third section has the relevant study and fourth section throws light on research gaps. CHAPTER 3 presents the overview of methodology proposed. CAHPTER 4 has the experimentation details and results. This section provides the details of dataset and explanation of training and testing of algorithm. Experiments were performed on three types of features that are text, speakers profile and the combination of both. It also detailed the results and comparison of result with the benchmark. CHAPTER 5 finally concludes the work and mentions future work for the research.

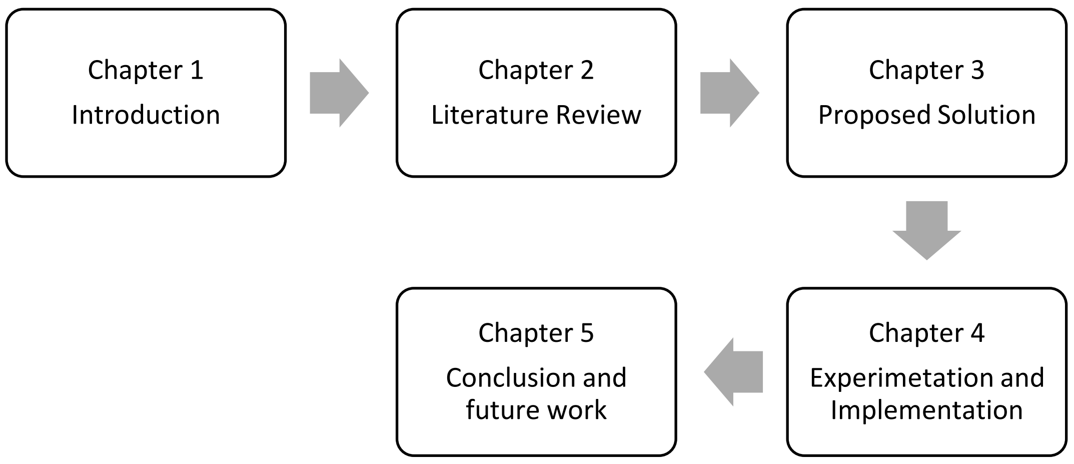

Figure 1.2: Thesis Organization

Chapter 2

Literature Review

CHAPTER 2: LITERATURE REVIEW

In this chapter, literature review is presented in detail. Section 2.1 contains the research methodology, Section 2.2 discusses the detailed analysis of selected papers, and approach followed, Section 2.3 presents the relative study along with their limitation.

2.1. Research Methodology

The research for literature review is performed in an organized way following the series of steps. Few categories are broadly defined to help in searching and sorting of articles. Along with that, selection and rejection criteria are described to develop a screening process for the papers. Quality assessment and data synthesis and extraction is also explained in upcoming sections.

2.1.1. Categories

Fake new detection mainly relies on news context and social context. Some useful categories of features are discussed below.

News context features

News context features describe the meta-information related to a piece of news that includes the source of news, a short title and the details of the story.

- *Source* of news are the author or publisher of the new article.
- *Short title* aims to catch viewers and describes the main topic of the Article.
- *Detail of the story* elaborates the news story and might shape the angle of source.

Based on these raw features, we can build different kinds of feature representations. Typically, the news content are linguistic [24-39] or visual [39, 40].

Linguistic-based: Linguistic approaches analyse the language pattern to extract cues of a deceptive message, following the belief that the language of truth differs from fabricated information [41, 42]. Fake news are often intentionally created to take advantage or to change societal behaviour rather than reporting facts. They often contain views and judgements of the source to confuse audience [43]. For example, fake news may contain some contradiction with facts or authentic resources [44]. Alternatively, a fake news item may contain more sentiment, negative emotion, less self-oriented pronouns, etc. [45] or a different grammatical structure [46]. Linguistic features are extracted from text from different levels that are word, sentence or document. Typically there are two types of linguistic features:

Lexical features: includes word level features, such a word count, frequency of positive or negative words [47] etc.

Syntactic words: includes sentence level features such a parts-of-speech tagging and frequency of function words etc.

Visual based: Visual-based features are extracted from images and videos. Real and fake images were classified based on various user and tweet related attributes using a classification framework [48]. In [40], the authors propose a set of numerous visual and statistical attributes to characterize different image distribution patterns while in [39] a combination of both visual as well as textual features of news item were used for fake news identification.

Social context features

Social context features are derived from user and social engagements of news on social network platforms. These features fall in three broad categories: user-based [32, 35, 36, 40, 49-51], based on generated posts [33-35, 50-53] and network-based [32, 40, 54-57].

User based: Capturing profiles of the users and characterising user-based features are the two handy operations for detecting fake news. The user-based features focus on the important noticeable features of the users who for the sake of news sharing interact with other users. The features can be segregated into individual level or group level. The individual level features shows trustworthiness and authenticity based on the users' gender, age, location, popularity or frequency of posts/tweets [58]. The group level features take into consideration the qualities and characteristics of all the users in totality [49].

Post based: Post-based features revolve around classifying useful information in order to know about the truthfulness of the news from numerous aspects of pertinent news. These features can be characterized as post level, group level, and temporal level. Post level features are exclusive features for posts that signify the reactions on social media, such as attitude, subject, and reliability. Group level features are intended at summing up the feature values for all pertinent posts [59]. Temporal level features deliberate upon the temporal differences of post level feature [60]. Some methods were proposed in the iterature to capture the variation in post over time [33, 61]. And based on these variation, various mathematical features can be calculated such as SpikeM parameters[62].

Network based: Network-based features can be extracted by making explicit networks of the users who published related social media posts. Different types of networks can be constructed. The stance network can be constructed with nodes showing all the pertinent news and the edge specifying the weights of resemblance [59, 63]. Co-occurrence network is another type of network, that shows the user engagements by counting whether relevant posts are written by the users to the same news articles[33]. Furthermore, the friendship network creates the network of users who post related news. Diffusion network is another extension of the friendship network, where users are represented by nodes and the information diffusion among them is represented by the edges [62].

2.1.2. Selection and Rejection Criteria

To establish a literature review in an organized way, some selection and rejection criteria have been developed. This criterion helps to ensure that selected research papers must fulfil these parameters to bring quality in work. All researches have been selected on the basis of following parameters:

1. ***Topic relevancy:*** article selection should be relevant to the subject under study. Only papers that are based on news classification are considered in the literature. Moreover, researches that go out of scope of news classification are excluded from study.
2. ***Year of publication:*** To make sure that the literature of this dissertation consists of state of the art techniques, papers only from the year 2014 – 2018 are inclusive. Whereas researches that are not from this time period are rejected.
3. ***Publishers:*** Publishers are also added to selection and rejection criteria. Papers from four famous scientific databases including IEEE, ACM, ELSEVIER, and SPRINGER are extracted. This rule is added to bring the authenticity and validity in the literature. Table 2.1 shows the details of different papers against their publishers.
4. ***Research impact:*** Another important factor while selecting paper is to check the impact of research in news classification. News classification with significant research and immense influence is considered in literature. Moreover, rest of the studies are excluded.
5. ***Experimentation and Facts:*** Researches with proven facts and supported by strong experimentation are considered in the literature. Researches and their results with hypothetical claim are not considered justified.
6. ***Repetition:*** finding that are based on significant and non-redundant studies are added in the literature review. Other redundant studies are rejected to be part of review.

Table 2.1: Research work per database

	Scientific Database	**Type**	**Selected Research Works**	**No. of Researches**
1.	IEEE	Journal	[40] [55]	2

		Conference	[28] [27] [54]	3
2.	ACM	Journal	[32]	1
		Conference	[33] [35] [39] [38, 53] [26, 52] [50] [37]	9
3.	ELSEVIER	Journal	[34] [57]	2
4.	SPRINGER	Journal	[24] [36, 56]	3
		Conference	[29] [51] [25]	3
5.	Others	Conference	[30] [31]	2
Total				25

2.1.3. Quality Assessment

In order to understand the important outcomes of selected research papers, we have developed quality criterion. It also defines the reliability of each selected research and its conclusive outcomes:

1. The data evaluation of the research is based on the facts and academic understanding without any vague statements.
2. The research is validated through proper validation methods.
3. The aim is to include most recent researches, because of our intention to investigate latest fake news detection techniques (Figure 2.1).
4. Originality of the research is another significant feature. Researches that are published in at least one of the four renowned and globally accepted scientific databases i.e. IEEE, SPRINGER, ELSEVIER and ACM are included. (Figure 2.2)

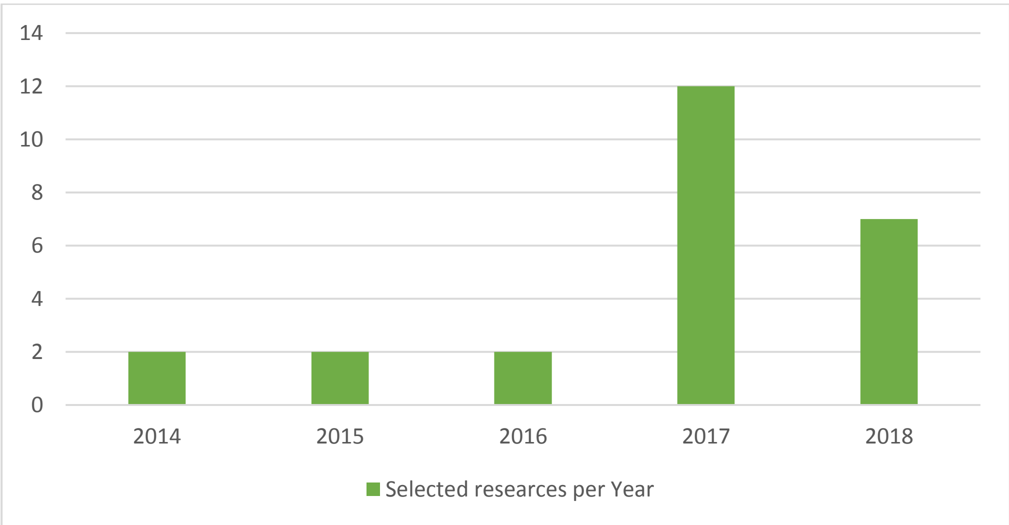

Figure 2.1: Detail of Selected Researches per year

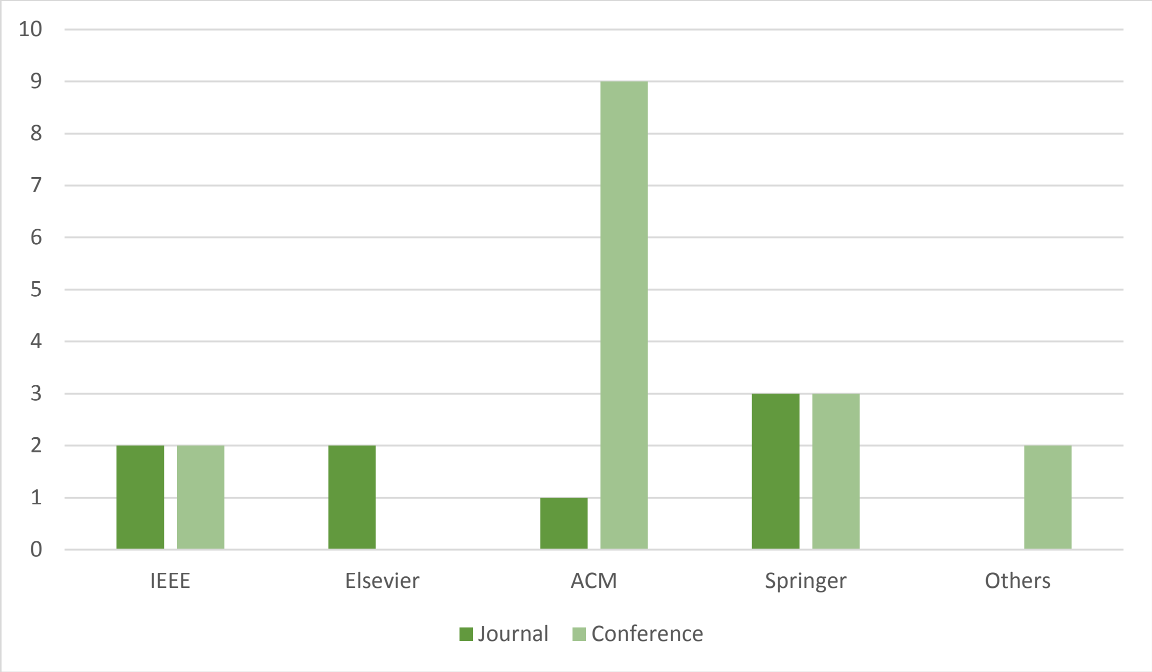

Figure 2.2: Selected Researches per publisher

2.1.4. Data Extraction and Synthesis

Table 2.2, shows the data synthesis and extraction is performed. Data extraction is done on our specific researches by extracting a significant amount of data from them according to our inclusion/exclusion criteria. While data synthesis is done by the detail study and analyse the processes and features used in our selected researches and put the papers in already discussed detection categories.

Table 2.2: Details of Data extraction and synthesis

Sr. #	**Descriptions**	**Details**
1.	Bibliographic information	topic, author, publication year, detail of publisher and the type of publication (conference or journal)
Data Extraction		
2.	Overview	Gist of paper and research targets
3.	Results	Results from the research
4.	Assumption(s)	For validating the results
5.	Validation	Validation approaches to prove the results.
Data Synthesis		
6.	Classification	Classification of new that are binary class and multiclass
7.	Features	Features used in the algorithm for classification

2.2. Analysis of Data

2.2.1. Classification of Data

This literature review is comprised of 25 research papers. In this section, these papers are studied placed into class label categories. This is demonstrated in Table 2.3.

Table 2.3: Classification of researches

Class Label	Researches
Binary Class	[57] [52] [26] [27] [28] [29] [34] [55] [56] [53] [51] [36] [24] [50] [40] [25] [32] [35] [39] [31] [33] [30]
Multiclass	[37] [31] [38]

2.2.2. Features for Classification

The elected papers use different sets of features for news classification. These studies are organized into 5 predefined categories of features after detailed analysis a shown in Table 2.4.

Table 2.4: Detail of Features for classification

Sr. #	Category	Features	Research Identification
1	News Content	Linguistic based	[24] [25] [26] [27] [28] [29] [30] [31] [32] [33] [34] [35] [36] [37-39]
2		Visual Based	[39] [40]
3	Social Context	User Based	[24] [50] [40] [32] [35] [51] [36]
4		Post Based	[53] [51] [52] [33] [34] [35] [50]
5		Network Based	[55] [56] [57] [32] [40]

The above table shows seven different categories that are defined from literature study. Analysing the papers, they are placed in the category they belong. It is observed that some papers combination of features from different categories.

2.3. Relevant Study

After analysing all the research data from different aspects, this section presents the methodologies and procedures used in study of news classification. Again, it comprises of papers that we have selected through the process of screening. Techniques that are proposed from 2014 to 2018 are discussed here. Anything that is not properly validated through experimentations and factual data is not added in this review. Special caution is taken to avoid redundancy and duplication. Table 2.5 shows the relevant study. It contains the title of the paper and description of the work done in that article.

Table 2.5: Selected studies for literature review

Sr. #	Title of the paper	Work done
1.	3HAN: A Deep Neural Network for Fake News Detection [25]	This paper works on a three level hierarchical method using deep learning. 3HAN is constructed in three levels that are a word, a headline and a sentence. It processes an article from bottom to top hierarchically. It gives attention weights to different parts of the article which can be visualized for further fact checking.
2.	Identification of Fake Reviews Using New Set of Lexical and Syntactic Features [26]	In this paper, the authors explore text classification of reviews through supervised learning algorithms. It provides a set of lexical and syntactic features and the focus is on the writing style.
3.	Fake news detection using naive Bayes classifier [27]	This paper proposes a method of text classification for fake new identification using naïve Bayes algorithm. The proposed approach was implemented and tested on the dataset containing Facebook posts.
4.	Implementation of Emotional Features on Satire Detection[28]	In this paper, the authors present a model for satire detection based on emotional features extracted from text. The model is implemented using supervised and unsupervised weighing approach with ensemble bagging.
5.	Detection of Online Fake News Using N-Gram Analysis and Machine	In this paper the authors investigate different feature extraction and machine learning techniques for text classification with N-Gram

	Learning Techniques [29]	analysis. It is concluded that the combination of Term Frequency-Inverted Document Frequency (TF-IDF) and linear support vector machine produces the best performance.
6.	Fake News Detection using Stacked Ensemble of Classifiers [30]	The authors propose a model for stance detection which has a stacked ensemble of classifiers.
7.	From Clickbait to Fake Detection: An Approach based on Detecting the Stan Headlines to Articles [31]	Clickbait is an attention grabbing headline to make the audience click on the links of news. The paper proposed a model to detect related and unrelated headlines of the article.
8.	Rumor Gauge: Predicting the Veracity of Rumors on Twitter [32]	This paper presents a solution to predict the veracity of rumours on social media. Rumour gauge is developed by identifying the features of news by analysing different aspects of information that are linguistic styles, users involved and the properties of user network. The veracity of rumours is generated by using hidden Markov models.
9.	CSI: A Hybrid Deep Model for Fake News Detection [33]	CSI model combines three characteristics of news (i.e. text, user response and the user promoting it) in three modules i.e. capture, score and integrate. First module captures the temporal pattern of users' response using recurrent neural networks. Second module analyses the behaviour of the source and these two modules are integrated in the third module to classify an article.
10.	The diffusion of misinformation on social media: Temporal pattern, message, and source [34]	This study analyses the misinformation spread on based on time series, text and the source of fake news. It was observed that on social media misinformation tends to spread wider and faster after publication as compared to true

		information. It was observed that the old news sometimes reappeared and gained visibility by users.
11.	Learning to Detect Misleading Content on Twitter [35]	Multimedia content is another source to mislead the audience. This paper targets the problem of multimedia content containing misinformation on twitter in real time. It proposed a new semi supervised learning approach with a set of features that generalizes the detection capabilities of trained models. It helps in detection even when the content is unseen and different from the training.
12.	Towards automated real-time detection of misinformation on Twitter [54]	It defines misinformation as a piece of news circulating on social media that conflicts a credible source. Credibility establishment is based on the premise that verified accounts share more credible information as compared to unverified accounts. The detection of misinformation is based on the mismatch ratio between tweets from verified and unverified users.
13.	FluxFlow: Visual Analysis of Anomalous Information Spreading on Social Media [55]	It presents a visual analysis system for analysing unusual spreading of information on social media. The system FluxFlow combines different advanced machine learning algorithms for the detection of abnormalities and presents a visualization pattern for deep analysis.
14.	Detecting misinformation in online social networks using cognitive psychology [56]	The work of this paper explores the cognitive psychology for evaluating the diffusion of misinformation. The cognitive process is based on four questions; credibility of source, consistency, coherency and general acceptability of a message. The proposed method combines

		the filtration of social media to measure the credibility of source and quality of news.
15.	A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis [57]	The trend of fake news arose in US presidential elections 2016. This study examined the root content, original source and evolution pattern of the news. It was observed that fake news are generated by unverified users and contains a link to a non-credible resource. The evolution pattern of fake and real news are different. From evolution pattern, it is analysed that the real news spread faster and wider but content of fake news changes with time.
16.	Hoaxy: A Platform for Tracking Online Misinformation [53]	The platform Hoaxy was introduced for the collection, detection and analysis of misinformation. It discovered that real news lags the misinformation by 10-20 hours. Fake news are popular among active users while facts are mostly grass-root activity.
17.	A Model for Identifying Misinformation in Online Social Networks [51]	A model that can identify suspicious behavioural patterns from users and news attributes is discussed in this paper. The model timely identifies and limits the diffusion of fake news.
18.	Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation [52]	This paper experiments by allowing a social media user to flag stories as fake, and if the story gets enough flags then it would be forwarded to a third party for fact checking. If the third party finds it fake, it will be marked as disputed. In short, the group of social media users can participate in prevention of fake news.

19.	Detection and visualization of misleading content on Twitter [36]	The system proposed extracts features from a source and its tweets. It trains the model in two steps using a semi supervised learning approach. The final output uses the agreement between the two models for the classification of new posts as guidance.
20.	Ranking-based Method for News Stance Detection [37]	Ranking based method was proposed to improve the performance of stance detection. The four labels are agree, discuss, disagree, and unrelated. According to this research there is no clear margins between two stances. Ranking based method maximizes the difference between true stance and false stance.
21.	Combining Neural, Statistical and External Features for Fake News Stance Identification [38]	This paper works on different kinds of features for stance detection. The authors compute a statistical feature from text, neural features from a deep recurrent model and external features from features engineering heuristics. At the end, deep neural layer combines the features for stance classification.
22.	EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection [39]	A neural network based approach to detection of fake news about newly emerging events is present. It derives visual and textual features that are invariant with respect to the event.
23.	Verifying information with multimedia content on twitter [24]	A comparative analysis of three methods of information verification for twitter is presented. These include textual patterns for truthfulness, topic-based veracity coherence, and a semi supervised learning technique based on two different verification classifiers. A technique that combines these methods is also proposed which is described to provide better authentication.

24.	Fake News Detection in Social Networks via Crowd Signals [50]	A detection algorithm based on Bayesian inference is proposed that uses crowd based labels on news items. The algorithm adapts based on the labelling accuracy of the crowd over time.
25.	Novel Visual and Statistical Image Features for Microblogs News Verification [40]	The content of visual information associated with a news item is used for its verification. Different visual and statistical features are used to classify a news item based on its image distribution pattern.

2.4. Research Gap

The preceding review of existing literature indicates that the fake news identification problem is often considered as a binary classification problem. However, most news cannot be explicitly classified as absolutely true or absolutely false. As argued earlier, they are generally a combination of misinformation and true news. Therefore, we believe a multi-class classification approach is more realistic. Multiclass classification has been used for stance detection where the authors' opinion or perception about a news item is automatically detected. The classification in this case predicts labels related to whether the author 'agrees', 'disagrees', 'discusses' a piece or the author's opinion is 'unrelated' [37, 38]. Moreover, most of existing related work has been targeted at identifying false information in the domain of social media. Thirdly, classification using the speaker's profile is in early phase of research and we believe the speaker's profile can be effectively used for fake news detection. In this research, we propose a novel multilevel approach for multi-label classification of fake news.

Chapter 3

Proposed Methodology

CHAPTER 3: PROPOSED METHODOLOGY

In this chapter, proposed methodology of the thesis has been discussed. Section 3.1 presents fake news detection using multi-layer supervised learning while Section 3.2 demonstrates an example of the proposed method being applied.

3.1. Fake News Detection Using Multi-Layer Supervised Learning

We propose a novel approach for fake news classification based on a three step process consisting of feature selection/extraction, relabelling and learning. Two of these steps are performed iteratively at multiple levels. Let $N = \{n_1, n_2, \ldots, n_n\}$ be a set of news items, L_0 be the set of original labels and $M = \{m_1, m_2, \ldots, m_n\}$ be the set of trained models. The details of terminology used are discussed in Table 3.1.

Table 3.1: Detail of terminologies used

Term	Purpose
$N = \{n_1, n_2, \ldots, n_n\}$	Set of News
L_0	Set of Original Labels
$M = \{m_1, m_2, \ldots, m_n\}$	Set of trained model.

An overview of the solution is shown in Figure 3.1.

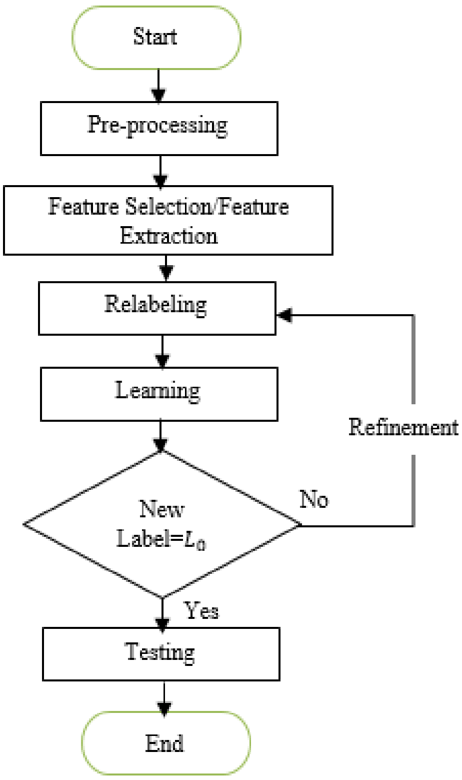

Figure 3.1: Fake News Detection Using Multi-Layer Supervised Learning

3.1.1. Data Pre-processing

In our approach, the dataset N is pre-processed before it can be used for training a classifier. This is achieved by first cleaning the dataset manually. Moreover, the metadata is converted to numeric values as some classifiers are designed solely for numerical data.

Text Pre-processing

Certain alterations like stop-word removal, tokenization, a lower casing, sentence breakdown, and punctuation removal were performed on the news before classification. This reduced the size of data by removing the unimportant information. Therefore a basic processing function to remove punctuation for each news was created.

Removing stop words: In any language, stop words are common words which are used to connect sentences and have no real meaning of their own. They increase the size of data and do not contribute when used in text classification. After removing these words from the news, the output were stored as tokens and passed on to the next step.

Stemming: The process that changes words into their root words is called stemming, and is used to decrease the number of words. For example, the words "write", "wrote" and "writer" were reduced to the word "writ". Stemming is used to make classifications quicker, efficient and less complex. Due to its accuracy, we use Porter stemmer, a commonly used algorithm for stemming.

3.1.2. Feature Selection and Feature Extraction

Feature Selection

"Feature selection is the process of selecting a subset of relevant features (attributes such as columns in tabular data) for use in model construction". We performed feature selection on metadata to exclude certain irrelevant features that may increase complexity and degrade the performance and/or accuracy of the algorithm.

Feature Extraction

Learning from high dimensional data is one of the difficulties of text categorization. There are a large number of words in news that leads to a high computing complexity while training a classifier. In addition to this, unrelated and redundant features can degrade the accuracy and performance of the classifiers. Therefore, the best course of action is to perform feature reduction by reducing the text feature size and avoiding a large feature space dimension. We calculate TFIDF for feature extraction.

Term Frequency-Inverted Document Frequency: The Term Frequency-Inverted Document Frequency (TF-IDF) is a weighting metric mostly used in information retrieval for text classification. It is a metric that measures the importance of a term in a document in the dataset. The more the term is used in the document the more significant that specific term is. However, this is countered by the frequency of the word in the corpus.

IDF is used to diminish the weight of the term frequency of the word that are unimportant and occur very frequently. For example, words such as “the” appear frequently in text, IDF reduces the impact of these terms.

3.1.3. Relabelling Process

In the relabelling step, we simplify the multiclass label problem by relabelling records. Initially, multiple class labels with similar properties are considered as a single class label. For example, for a set of original labels $L_0 = \{1,2,3,4,5\}$, the first three labels i.e. 1, 2 and 3 are considered as a single label. On the other hand, 4 and 5 show higher rating so they are high rating labels. Consequently, we have reduced the multiclass labels into 2 class labels as shown in Figure 3.2. From here on, we solve the multiclass problem with binary class solutions.

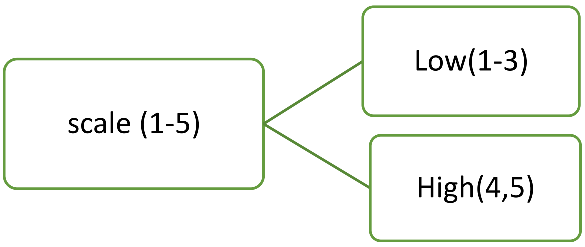

Figure 3.2: Example of relabelling process

3.1.4. Learning

After the relabelling process, we train classifiers on the new class labels assigned in the previous step. The learning algorithm determines patterns in the training data that enable it to map input data attributes to the new class labels. The resulting ML model $m_i \in M$ captures these relationships.

For metadata, we train support vector machine (SVM). For the experimentation of text and combination of features, we propose decision tree.

SVM

A support vector machine is used to determine the pose of a hyperplane in an n-dimensional feature space such that each data point is distinctly classified. Since many such planes might exist, the intended unique solution is the one which has the maximum distance from data points of each of the two classes. This provides the maximum margin for data points that may arise in future resulting in better confidence in their classification.

Decision Tree

A graphical representation of all classification paths that may be taken to reach a decision is known as a decision tree. It starts at a single root node then branches off based on each possible condition. The different branches of the resulting tree represent the possible solutions.

A decision tree is a tree where each node represents a feature (attribute), each link (branch) represents a decision (rule) and each leaf represents an outcome (categorical or continues value).

3.1.5. Refinement

We define the process of refinement as the process of relabelling and (re-)learning iteratively. Refinement is performed for each binary class individually which means that labels in each class are refined such that they move one hierarchical level closer to the original labels L_o. Consider the previous example, the class label *low* is relabelled again as *low* (1, 2) and *high* (3). This relabelled data is then used to train a new model $m_j \in M$.

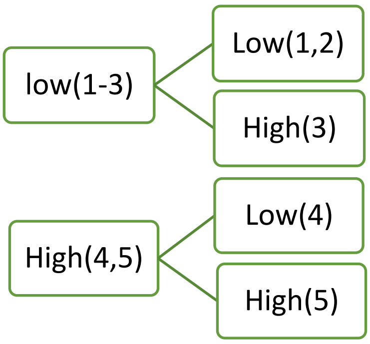

Figure 3.3: Example of refinement

Label refinement repeats itself until new labels converge to the original labels. The result is multiple trained models $M = \{m_1, m_2, \dots, m_n\}$ that will be used for classification. For the previous example, different trained models are represented by green shaded area in Figure 3.4.

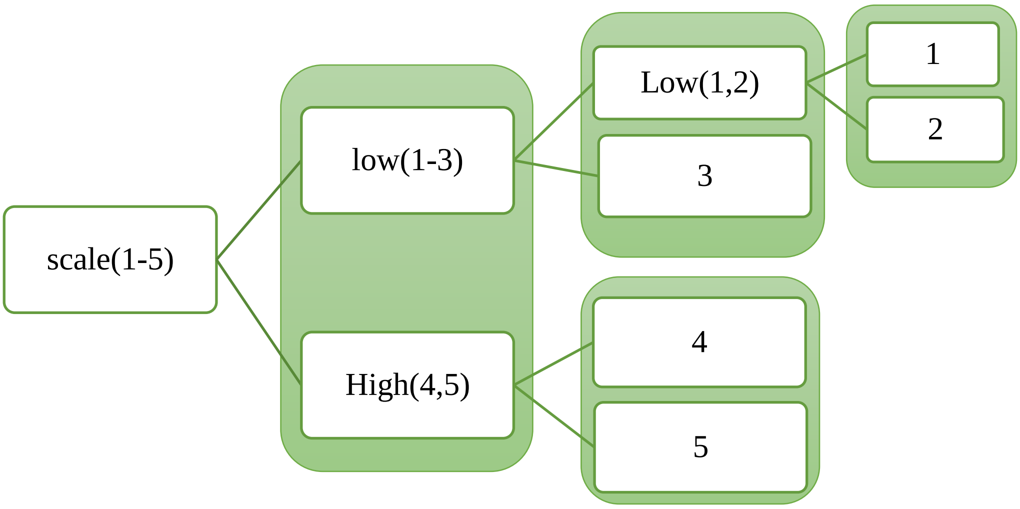

Figure 3.4: Example of proposed methodology

3.2. Classification Example using Proposed Solution

Continuing with the previously discussed example, a record is processed with the proposed algorithm and the trained models are m_1, m_2, m_3 and m_4. The system classifies the input using four models one by one. Initially the input is classified as *low* and *high* using model m_1. Suppose the first model produces output of *low*, which will became the input of m_2. And if the output is *low*, we pass the input to model m_4. It may produce the result of *low* (1) or *high* (2) as final result.

Similarly, if m_1 produces output of *high* (4, 5), and m_3 will be executed. It can generate output of *low* (4) or *high* (5) as final output.

This model is capable of generating outputs in following sequence:

Table 3.2: processing sequences

	Process	Output
	$m_1 \rightarrow m_2 \rightarrow m_4$	1
	$m_1 \rightarrow m_2 \rightarrow m_4$	2
	$m_1 \rightarrow m_2$	3
	$m_1 \rightarrow m_3$	4
	$m_1 \rightarrow m_3$	5

Chapter 4

Experimentation and Results

CHAPTER 4: EXPERIMENTATION AND RESULTS

This chapter presents the experimentation details for this research. In section 4.1, the dataset used is discussed. Section 4.2 is related to experiments on textual portion of dataset. Section 4.3 gives details of meta-data classification. In Section 4.4 the trained models are tested by performing different testing techniques i.e. holdout method, testing on test set and cross validation. Finally, Section 4.5 has the detail of results and comparison with the benchmark method.

4.1. Dataset and Data Pre-processing

The proposed solution has been evaluated using the LIAR dataset [64]. The dataset contains 12,836 short statements in POLITIFACT.COM[3]. The training dataset contain 10,269 records and testing set contains 1283 records. There are 14 attributes that belongs to news as well as social context including a short statement and the speaker's profile.

Speaker's profile contains speaker name, designation, party affiliation, state, context/venue of speech, and credit history. The credit history consist of the historical records of statements for each speaker. The six class labels are L_o = *pants-fire, false, barely-true, half-true, mostly-true,* and *true*. After manually cleaning the dataset, a total of 10235 records remain which are used for experimentation.

4.2. Metadata Classification

Metadata features include speaker's profile and context of news.

4.2.1. Pre-processing

The dataset was then codified into numeric dataset manually. The main reasons of manual codification is the spelling mistakes and use of acronyms in the dataset. For example, '*fb*' is used for '*Facebook*' in some records.

4.2.2. Feature Selection

For this purpose, forward feature selection technique is used. It is an iterative method in which we start with a set containing zero features. In every iteration we keep adding the

[3] https://www.politifact.com/

feature, from among other features, which best improves our model until adding new feature does not provide better performance.

The feature selection is performed using Weka 3.8 and 5 features were extracted that are *barely-true* count, *false* count, *half-true* count, *mostly-true* count and *pants on fire* count. It is clear that all the selected features belongs to speaker's credit history. The Weka summary of feature selection is shown in Figure 4.1.

```
=== Attribute Selection on all input data ===

Search Method:
        Greedy Stepwise (forwards).
        Start set: no attributes
        Merit of best subset found:    0.079

Attribute Subset Evaluator (supervised, Class (nominal): 1 label):
        CFS Subset Evaluator
        Including locally predictive attributes

Selected attributes: 6,7,8,9,10 : 5
                     barely true counts
                     false counts
                     half true counts
                     mostly true counts
                     pants on fire counts
```

Figure 4.1: Weka Summary of Feature Selection

The column chart in Figure 4.2 shows the accuracies of different sets of features.

Figure 4.2: Accuracies of different sets of features

4.2.3. Relabelling Dataset

As already discussed, the dataset has 6 labels, two of them belong to the false category and four belong to the true category. Initially, the dataset was relabelled in such a way that class labels {*false, pants on fire}* are considered false. Similarly, {*barely-true, half-true, mostly-true, true}* are considered as true as shown in Figure 4.3.

id	new label	original label	barely true counts	false counts	half true counts	mostly true counts	pants on fire counts
1.json	FALSE	pants-fire	0	0	0	0	1
100.json	TRUE	TRUE	12	5	9	3	4
1000.json	FALSE	pants-fire	8	22	6	4	16
10002.jsor	TRUE	barely-true	6	9	3	1	4
10003.jsor	TRUE	barely-true	1	0	1	0	0
10006.jsor	TRUE	barely-true	10	6	6	6	4
10007.jsor	FALSE	FALSE	11	3	8	6	5
10008.jsor	FALSE	FALSE	5	1	6	4	2
10009.jsor	FALSE	FALSE	5	5	11	8	3
1001.json	FALSE	FALSE	0	0	0	1	0
10010.jsor	TRUE	half-true	10	6	6	6	4
10012.jsor	TRUE	half-true	70	71	160	163	9
10013.jsor	TRUE	mostly-true	1	0	1	2	0
10016.jsor	FALSE	FALSE	7	6	3	5	1
10017.jsor	TRUE	half-true	2	2	4	2	0

Figure 4.3: Relabelled dataset

4.2.4. Learning

As discussed in the proposed methodology, the classifier is trained on new labels. At first, the machine is trained for *all trues* = {*barely-true, half-true, mostly-true, true}* and *all false*= {*false, pants-fire*} class labels.

4.2.5. Refinement Process

After m_1 is trained on new all trues and all false class labels, they are considered two different paths. The training dataset is relabelled to new labels one step closer to the original such as true is classified further as {*barely-true*, *half-true}* and {*mostly-true* and *true*}. The grouping of class labels is clear from the Figure 4.4.

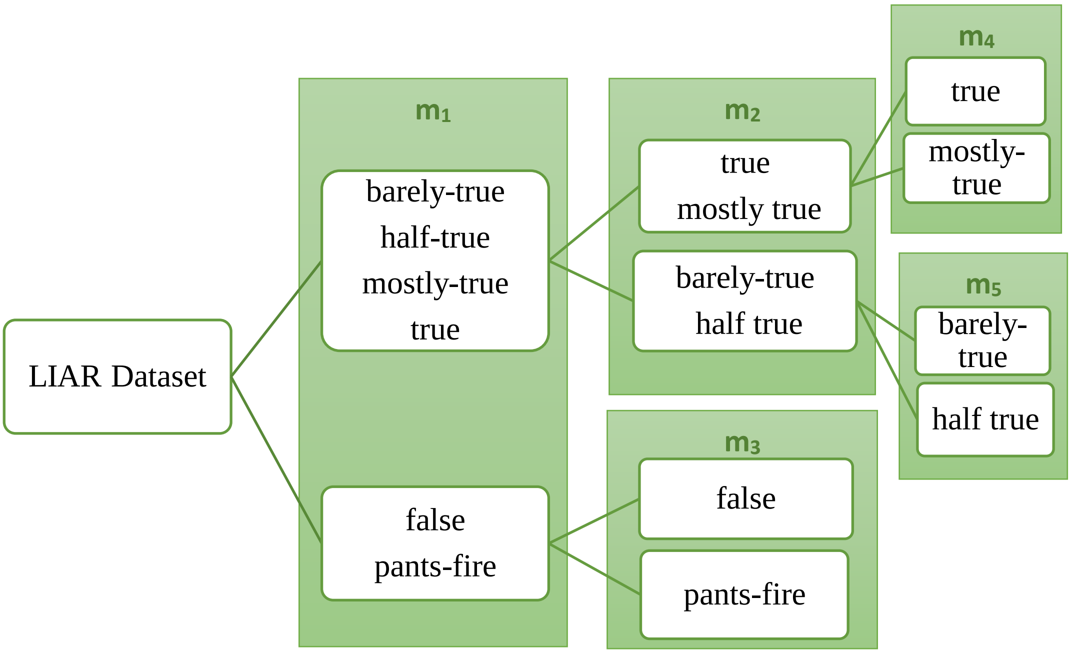

Figure 4.4: Refinement process of Experiments

The 5 models, m_1, m_2, m_3, m_4 and m_5, are trained with the following labels:

Table 4.1 Detail of labels of each model

Machine	**Label 1**	**Label 2**
m_1	*barely-true* *half-true* *mostly-true* *true*	*False* *Pants-fire*
m_2	*True* *Mostly-true*	*Barely-true* *Half-true*
m_3	*False*	*Pants-fire*
m_4	*True*	*Mostly-true*
m_5	*Half-true*	*Barely-true*

When input is given with its feature to the algorithm, it relabels the training dataset and trains the classifiers. The system classifies the input using five models one by one. The classification process will always start from m_1.

Initially the input is classified as all true and all fakes using model m_1. This model generates outputs in following sequence of machines:

Table 4.2: Sequence of machines to generate output

	Process	Output
1.	$m_1 \rightarrow m_2 \rightarrow m_4$	*True*
2.	$m_1 \rightarrow m_2 \rightarrow m_4$	*mostly-true*
3.	$m_1 \rightarrow m_2 \rightarrow m_5$	*Half-true*
4.	$m_1 \rightarrow m_2 \rightarrow m_5$	*Barely-true*
5.	$m_1 \rightarrow m_3$	*False*
6.	$m_1 \rightarrow m_3$	*Pants-fire*

4.3. Text Classification

Following activities are performed on textual attributes of the dataset for initial understanding and experiments. The text classification process is shown in Figure 4.5.

Pre-processing: The classification process includes pre-processing the text of news consists of tokenization, stop word removal and stemming.

Feature Extraction: Features extraction results in 7553 attributes.

Multi Layered Decision tree Learning: It is the iterative process of re-(labelling) and learning. We train multi-layered decision tree classifiers to predict the class of news. The labelling and learning process is same as for metadata.

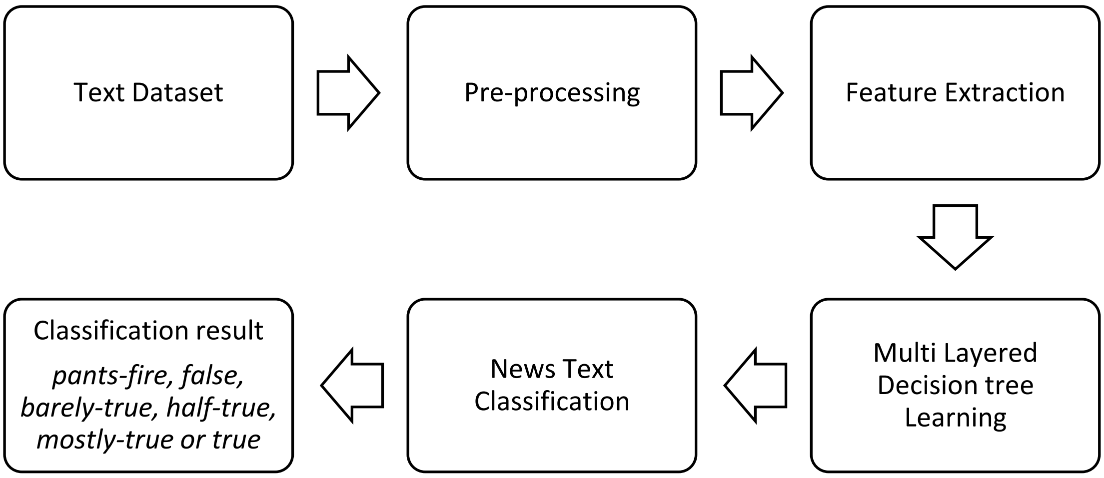

Figure 4.5: Text Classification Process

4.4. Testing

We perform three types of experiments i.e. experiments on metadata, text and a Hybrid experiment on metadata and text together.

4.1.1. Metadata Classification

The algorithm is tested using different classification algorithms that are SVM, decision tree and naïve Bayes, but only classification using SVM is illustrated in this section.

Holdout testing using SVM

For testing, the dataset is divided into training and test datasets in the ratio of 70:30. Thus, of the total 10235 records, 7165 were used for training and 3070 records were used for testing purposes. The detailed statistics of holdout testing are as follows:

1. After training, the model m_1 produces 85.57% accurate results for *all trues* and *all false* class labels. Among these 448 out of 811 records belong to the *false* label while 2180 of 2260 belong to the *true* category.
2. The accuracy of m_2 is 80.59% which means 1757 out of 2180 records were classified correctly. Among these 1099 records belong to {*mostly-true, true*} class label and 1081 records belong to {*barely-true, half-true*} class label.
3. The model m_3 produces an accuracy of 89.7% which means 402 out of the 448 records were predicted correctly. 308 *false* records and 94 *pants-fire* records are predicted correctly.
4. The accuracy of m_4 is 80.7% i.e. 877 records out of 1099 records are accurately classified. 584 of these records belong to *mostly-true* and 303 belong to the *true* class label.
5. Similarly, m_5 produces 85.8% accuracy, which means 870 records are correctly predicted. Among these 307 records are *barely-true* and 440 records are *half-true*.

Collectively, 2036 records are correctly predicted out of 3070 records, hence the accuracy of the proposed solution with hold-out testing method is 66.29%.

Testing on test set using SVM

The complete dataset of size 10235 is used for training and a separately given testing dataset is used for testing which consists of 1265 records.

The detailed statistics are as follows:

1. After training the m_1 produces 78.8% accurate results for *all trues* and *all false* class labels. Among these, 130 records were *false* while 867 records were *true*.
2. Accuracy of m_2 is 65.7%, 292 of these records belong to {*mostly-true, true}* class label and 278 records belongs to {*barely-true, half-true}* class label.
3. The model m_3 has the accuracy of 82.30%. Among these, 67 *false* records and 40 *pants-fire* records are predicted correctly.
4. The accuracy of m_4 is 68.15% that are 149 *mostly-true* records and 60 true records.
5. Similarly, m_5 produces 69.7% accuracy, which includes 75 *barely-true* records and 119 *half-true* records

Collectively, 500 records are correctly predicted out of 1256 records, hence the accuracy of proposed solution on testing data is 39.5%.

Cross Validation using SVM

The complete dataset of size 10235 is used for k fold classification where $k = 10$.

The detailed statistics of holdout testing are as follows:

1. After training the m_1 produces 80% accurate results for *all trues* and *all false* class labels. Out of which, 1207 records were false and 6981 records were true.
2. Accuracy of m_2 is 69.08%, 2408 of these records belong to *mostly-true and true* class label and 2415 records belongs to *barely-true and half-true* class label.
3. The model m_3 has the accuracy of 82.02%, 687 *false* records and 303 *pants-fire* records are predicted correctly.
4. The accuracy of m_4 is 70.39% i.e. 1247 *mostly-true* records and 448 true records.
5. Similarly, m_5 produces 72.38% accuracy, which includes 657 *barely-true* records and 1091 *half-true* records

Collectively, 4433 records are correctly predicted out of 10235 records. Hence the accuracy of proposed method on testing data is 43.3%.

Results

Decision tree and SVM classifier produces better results. Percentage accuracy of the experiments are detailed in the Table 4.3. In the Table 4.3, the term single level refers to training a model to all the six class labels at once. The accuracy peaks of each method are shown in bold.

Table 4.3: Experimentation results of different testing methods for metadata features

	Holdout method	Test set method	Cross validation
Single Level SVM	22.5	21.10	21.35
Single Level Decision tree	40.6	38.6	39.35
Proposed method with Decision tree	60.4	**39.7**	**43.7**
Proposed solution with SVM	**66.29**	39.5	43.38

According to these results, the proposed solution with SVM classifier in holdout method gives higher accuracy than other combinations.

4.1.2. Text Classification

The algorithm is tested using different classification algorithms that are decision tree and random forest, but only classification using decision tree is illustrated in this section.

Holdout Method using decision tree

For testing, the dataset is divided into training and test datasets in the ratio of 70:30. Thus, of the total 10235 records, 7165 were used for training and 3070 records were used for testing purposes. The detailed statistics of holdout testing are as follows:

1. After training the m_1 produces 70.3% accurate results for *all trues* and *all false* class labels which means 2158 records were classified correctly.129 belong to false category and 2029 belongs to true category.
2. The accuracy of m_2 is 54%.
3. The model m_3 produces 491 correctly predicted *false* record and 27 correctly predicted *pants-fire.*
4. The accuracy of m_4 produces 210 *mostly-true* and 115 *true* records.
5. Similarly, m_5 produces 211 *half-true* and 107 *barely-true* records.

Collectively, 1161 records are correctly predicted out of 3070 records, hence the accuracy of the proposed solution with hold-out testing method is 37.8%.

Test set method using Decision Tree

The complete dataset of size 10235 is used for training and a separately given testing dataset is used for testing which consists of 1265 records.

The detailed statistics are as follows:

1. After training, the m_1 produces 68.9% accurate results for *all trues* and *all false* class labels. Among these, 52 records were *false* while 821 records were *true*.
2. Accuracy of m_2 is 53%, 208 of these records belong to {*mostly-true, true}* class label and 231 records belongs to {*barely-true, half-true}* class label.
3. The model m_3 has the accuracy of 67%. Among these, 30 *false* records and 5 *pants-fire* records are predicted correctly.
4. The m_4 produces 69 *mostly-true* records and 43 true records.
5. Similarly, m_5 produces 43 *barely-true* records and 67 *half-true* records.

Collectively, 257 records are correctly predicted out of 1256 records, hence the accuracy of the proposed solution on testing data is 20.3%.

Results

It has been observed that the decision-tree classifier produces better results. Percentage accuracy of the experiments is detailed in the Table 4.4 where the accuracy peaks of each method are shown in bold.

Table 4.4: Experimentation results of different testing methods for text classification

	Holdout method	Test set method
Proposed method with Decision tree	**37.8**	20.28
Proposed solution with random forest	22.5	**21.9**

According to these results, the proposed solution with decision tree classifier in holdout method gives the higher accuracy than other combinations.

4.1.3. Hybrid Classification

In this section the combination of all features that are TF-IDF matrix and metadata are classified using proposed solution.

Holdout method

For testing, the dataset is divided into training and test datasets in the ratio of 70:30. Thus, of the total 10235 records, 7165 were used for training and 3070 records were used for testing purposes. The detailed statistics of holdout testing are as follows:

6. After training the m_1 produces 77.7% accurate results for *all trues* and *all false* class labels.
7. The accuracy of m_2 is 67.06%.
8. The model m_3 produces an accuracy of 80.9% 228 *false* records and 94 *pants-fire* records are predicted correctly.
9. The accuracy of m_4 is 68.6%. 301 of these records belong to *mostly-true* and 164 belong to the *true* class label.
10. Similarly, m_5 produces 75.6% accuracy. Among these 212 records are *barely-true* and 285 records are *half-true*.

Collectively, 1284 records are correctly predicted out of 3070 records, hence the accuracy of the proposed solution with hold-out testing method is 41.8%.

Test set Method

The complete dataset of size 10235 is used for training and a separately given testing dataset is used for testing which consists of 1265 records.

The detailed statistics are as follows:

1. After training, the m_1 produces 76.6% accurate results for *all trues* and *all false* class labels. Among these, 170 records were *false* while 799 records were *true*.
2. Accuracy of m_2 is 63.2%, 257 of these records belong to {*mostly-true, true}* class label and 248 records belongs to {*barely-true, half-true}* class label.
3. The model m_3 has the accuracy of 85.2%. Among these, 100 *false* records and 45 *pants-fire* records are predicted correctly.
4. The m_4 produces 66.9% results, 114 *mostly-true* records and 58 true records are predicted correctly.
5. Similarly, m_5 produces 74% results that are 80 *barely-true* records and 104 *half-true* records.

Collectively, 501 records are correctly predicted out of 1256 records, hence the accuracy of proposed solution on testing data is 39.6%.

Results

It was observed that the decision-tree classifier produces better results. Percentage accuracy of the experiments are detailed in the Table 4.5 where the accuracy peaks of the method are shown in bold.

Table 4.5: Experimentation results of different testing methods for hybrid method

	Holdout method	Test set method
Proposed method with Decision tree	**41.8**	39.6

According to these results, the proposed solution with decision-tree classifier in holdout method gives higher accuracy than other combinations.

4.2. Analysis of Results

Results of all the three types of experiments are combined in the Table 4.6. We achieved our accuracy peak in holdout method with credit history features and SVM classifier. For test set method we achieved approximately similar accuracy for credit history as well as for the hybrid features with decision tree.

Table 4.6: Combined Results of all experiments

	Features	Holdout method	Test set method	Cross validation
Single Level SVM	Credit History	22.5	21.10	21.35
Single Level Decision Tree	Credit History	40.6	38.6	39.35
Proposed method with Decision tree	Credit History	60.4	**39.7**	**43.7**
Proposed solution with SVM	Credit History	**66.29**	39.5	43.38
Proposed method with Decision Tree	Text	37.8	20.28	-
Proposed method with Decision Tree	Text+ all features	41.8	39.6	-

4.3. Comparison with benchmark

The problem of fake news detection has been explored by Wang et. al [65] where the authors classify text as well as the corresponding metadata based on a hybrid Convolutional Neural Networks framework that integrates text and meta-data. The authors have only used the test dataset for evaluation of their approach and not considered the hold-out or cross validation methods. Their analysis shows that best results are obtained when using all attributes including text.

Table 4.7: Comparison with benchmark

Model	**Features**	**Test accuracy %**
Wang's Hybrid CNN [64]	Text+ speakers profile	27.4
Proposed Solution with SVM	Credit history	39.5
Proposed Solution with Decision Tree	Credit history	**39.7**
Proposed solution with Random Forest	Text features	21.9
Proposed Solution	Text + all features	39.6

The proposed solution outperforms the benchmark with accuracy of 39.7.

It is also observed from the experimentation that speaker's credit history contributes the most in fake news detection.

Chapter 5

Conclusion and Future Work

CHAPTER 5: CONCLUSION AND FUTURE WORK

Fake news is a growing concern worldwide and automatic identification schemes are fast becoming a necessity due to the decentralized nature of the interconnected digital world. Detection of fake news is often considered a binary classification problem but most news items are a combination of both, accurate and false information cannot be explicitly classified as absolutely true or absolutely false. A multi-class classification approach similar to the one used for an author's stance detection can be applied for identify where a news item lies on the spectrum from absolutely fake to completely accurate. The use of a speaker's profile for such classification needs to be evaluated for communication means beyond social media.

We propose a multi-level fake news detection method based on supervised learning. The idea is to create a hierarchy of labels such that at each node of each level in the hierarchy is a binary classification problem. The input dataset consists of multiple labels and the speakers' profile which includes the speaker's name, designation, party affiliation, credit history, etc. The six labels used are pants-fire, false, barely-true, half-true, mostly true, and true. The data is first pre-processed to make it suitable for multi-level classification. The three steps of the proposed approach are then applied to the processed data. Firstly, feature selection/extraction is used to determine the features that best span the variation in the data. Secondly, the six labels are converted to binary labels in a process we call relabelling. Lastly, in the learning step classifiers are trained on the relabelled data. Data elements in each of the two labels of the previous iteration are further relabelled to the next lower level of the hierarchy. The learning step is repeated for the relabelled data. Thus, the 'relabelling' and 'learning' steps are iteratively performed until the lowest level of the label hierarchy is classified by the machine learning algorithms. Three modes of classification are used i.e. metadata based classification, news text classification and a combination of both. In the first mode, forward feature selection is used which selects features related to credit history. In news text classification feature extraction is used which produces a TF-IDF matrix. In the third mode, both these modes are used together. The classifiers used in this research are support vector machines and decision-tree classifiers. Different methods of evaluation of the classification accuracy are used including hold-out method, separate test data, and cross-validation. The results are compared with benchmark techniques and the results demonstrate higher accuracy as compared to existing techniques. The proposed method outperforms the benchmark with an accuracy of 39.5%. Our experiments indicate that the profile of the source of information contributes significantly to fake news detection.

The proposed method can also be tested on other multi class problems. Another direction of future research is to use ensembles instead of simple classifiers.

References

[1] H. Rashkin, E. Choi, J. Y. Jang, S. Volkova, and Y. Choi, "Truth of varying shades: Analyzing language in fake news and political fact-checking," in *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, 2017, pp. 2931-2937.

[2] S. Vosoughi, D. Roy, and S. Aral, "The spread of true and false news online," *Science,* vol. 359, pp. 1146-1151, 2018.

[3] E. S. Reed, E. Turiel, and T. Brown, "Naive realism in everyday life: Implications for social conflict and misunderstanding," in *Values and knowledge*, ed: Psychology Press, 2013, pp. 113-146.

[4] C. Metzgar, "Confirmation Bias: A Ubiquitous Phenomenon in Many Guises," *Professional Safety,* vol. 58, p. 44, 2013.

[5] D. Kahneman and A. Tversky, "Prospect theory: An analysis of decision under risk," in *Handbook of the fundamentals of financial decision making: Part I*, ed: World Scientific, 2013, pp. 99-127.

[6] A. Tversky and D. Kahneman, "Advances in prospect theory: cumulative representation of uncertainty," in *Readings in Formal Epistemology*, ed: Springer, 2016, pp. 493-519.

[7] H. Tajfel and J. C. Turner, "An integrative theory of intergroup conflict," *The social psychology of intergroup relations,* vol. 33, p. 74, 1979.

[8] H. Tafjel and J. C. Turner, "The social identity theory of intergroup behavior," *Psychology of intergroup relations,* pp. 7-24, 1986.

[9] S. E. Asch and H. Guetzkow, "Effects of group pressure upon the modification and distortion of judgments," *Groups, leadership, and men,* pp. 222-236, 1951.

[10] J.-N. Kapferer, *Rumors: Uses, Interpretation and Necessity*: Routledge, 2017.

[11] Z. Jin, J. Cao, Y.-G. Jiang, and Y. Zhang, "News credibility evaluation on microblog with a hierarchical propagation model," in *Data Mining (ICDM), 2014 IEEE International Conference on*, 2014, pp. 230-239.

[12] H. Shen, F. Ma, X. Zhang, L. Zong, X. Liu, and W. Liang, "Discovering social spammers from multiple views," *Neurocomputing,* vol. 225, pp. 49-57, 2017.

[13] Y. Chen, N. J. Conroy, and V. L. Rubin, "News in an online world: The need for an “automatic crap detector”," *Proceedings of the Association for Information Science and Technology,* vol. 52, pp. 1-4, 2015.

[14] V. L. Rubin, "Deception detection and rumor debunking for social media," *The SAGE Handbook of Social Media Research Methods,* pp. 342-363, 2017.

[15] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, "Fake news detection on social media: A data mining perspective," *ACM SIGKDD Explorations Newsletter,* vol. 19, pp. 22-36, 2017.

[16] C. Andrieu, N. De Freitas, A. Doucet, and M. I. Jordan, "An introduction to MCMC for machine learning," *Machine learning,* vol. 50, pp. 5-43, 2003.

[17] N. M. Nasrabadi, "Pattern recognition and machine learning," *Journal of electronic imaging,* vol. 16, p. 049901, 2007.

[18] C. Robert, "Machine learning, a probabilistic perspective," ed: Taylor & Francis, 2014.

[19] F. Sebastiani, "Machine learning in automated text categorization," *ACM computing surveys (CSUR),* vol. 34, pp. 1-47, 2002.

[20] C. E. Rasmussen, "Gaussian processes in machine learning," in *Advanced lectures on machine learning,* ed: Springer, 2004, pp. 63-71.

[21] D. Freitag, "Machine learning for information extraction in informal domains," *Machine learning,* vol. 39, pp. 169-202, 2000.

[22] B. Pang, L. Lee, and S. Vaithyanathan, "Thumbs up?: sentiment classification using machine learning techniques," in *Proceedings of the ACL-02 conference on Empirical methods in natural language processing-Volume 10*, 2002, pp. 79-86.

[23] M. Barreno, B. Nelson, A. D. Joseph, and J. D. Tygar, "The security of machine learning," *Machine Learning*, vol. 81, pp. 121-148, 2010.

[24] C. Boididou, S. E. Middleton, Z. Jin, S. Papadopoulos, D.-T. Dang-Nguyen, G. Boato, *et al.*, "Verifying information with multimedia content on twitter," *Multimedia Tools and Applications*, pp. 1-27, 2017.

[25] S. Singhania, N. Fernandez, and S. Rao, "3HAN: A Deep Neural Network for Fake News Detection," Cham, 2017, pp. 572-581.

[26] R. K. Dewang and A. K. Singh, "Identification of Fake Reviews Using New Set of Lexical and Syntactic Features," presented at the Proceedings of the Sixth International Conference on Computer and Communication Technology 2015, Allahabad, India, 2015.

[27] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in *Electrical and Computer Engineering (UKRCON), 2017 IEEE First Ukraine Conference on*, 2017, pp. 900-903.

[28] P. P. Thu and N. New, "Implementation of emotional features on satire detection," in *Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD), 2017 18th IEEE/ACIS International Conference on*, 2017, pp. 149-154.

[29] H. Ahmed, I. Traore, and S. Saad, "Detection of online fake news using N-gram analysis and machine learning techniques," in *International Conference on Intelligent, Secure, and Dependable Systems in Distributed and Cloud Environments*, 2017, pp. 127-138.

[30] J. Thorne, M. Chen, G. Myrianthous, J. Pu, X. Wang, and A. Vlachos, "Fake news stance detection using stacked ensemble of classifiers," in *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, 2017, pp. 80-83.

[31] P. Bourgonje, J. M. Schneider, and G. Rehm, "From clickbait to fake news detection: an approach based on detecting the stance of headlines to articles," in *Proceedings of the 2017 EMNLP Workshop: Natural Language Processing meets Journalism*, 2017, pp. 84-89.

[32] S. Vosoughi, M. N. Mohsenvand, and D. Roy, "Rumor gauge: Predicting the veracity of rumors on Twitter," *ACM Transactions on Knowledge Discovery from Data (TKDD)*, vol. 11, p. 50, 2017.

[33] N. Ruchansky, S. Seo, and Y. Liu, "Csi: A hybrid deep model for fake news detection," in *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, 2017, pp. 797-806.

[34] J. Shin, L. Jian, K. Driscoll, and F. Bar, "The diffusion of misinformation on social media: Temporal pattern, message, and source," *Computers in Human Behavior*, vol. 83, pp. 278-287, 2018.

[35] C. Boididou, S. Papadopoulos, L. Apostolidis, and Y. Kompatsiaris, "Learning to Detect Misleading Content on Twitter," presented at the Proceedings of the 2017 ACM on International Conference on Multimedia Retrieval, Bucharest, Romania, 2017.

[36] C. Boididou, S. Papadopoulos, M. Zampoglou, L. Apostolidis, O. Papadopoulou, and Y. Kompatsiaris, "Detection and visualization of misleading content on Twitter," *International Journal of Multimedia Information Retrieval*, vol. 7, pp. 71-86, 2018.

[37] Q. Zhang, E. Yilmaz, and S. Liang, "Ranking-based Method for News Stance Detection," presented at the Companion Proceedings of the The Web Conference 2018, Lyon, France, 2018.

[38] G. Bhatt, A. Sharma, S. Sharma, A. Nagpal, B. Raman, and A. Mittal, "Combining Neural, Statistical and External Features for Fake News Stance Identification," presented at the Companion Proceedings of the The Web Conference 2018, Lyon, France, 2018.

[39] Y. Wang, F. Ma, Z. Jin, Y. Yuan, G. Xun, K. Jha, *et al.*, "EANN: Event Adversarial Neural Networks for Multi-Modal Fake News Detection," presented at the Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, United Kingdom, 2018.

[40] Z. Jin, J. Cao, Y. Zhang, J. Zhou, and Q. Tian, "Novel visual and statistical image features for microblogs news verification," *IEEE transactions on multimedia*, vol. 19, pp. 598-608, 2017.

[41] J. Bachenko, E. Fitzpatrick, and M. Schonwetter, "Verification and implementation of language-based deception indicators in civil and criminal narratives," in *Proceedings of the 22nd International Conference on Computational Linguistics-Volume 1*, 2008, pp. 41-48.

[42] D. F. Larcker and A. A. Zakolyukina, "Detecting deceptive discussions in conference calls," *Journal of Accounting Research*, vol. 50, pp. 495-540, 2012.
[43] Y. Chen, N. J. Conroy, and V. L. Rubin, "Misleading online content: Recognizing clickbait as false news," in *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, 2015, pp. 15-19.
[44] V. W. Feng and G. Hirst, "Detecting deceptive opinions with profile compatibility," in *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 2013, pp. 338-346.
[45] Y. Long, Q. Lu, R. Xiang, M. Li, and C.-R. Huang, "Fake news detection through multi-perspective speaker profiles," in *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, 2017, pp. 252-256.
[46] B. D. Horne and S. Adali, "This just in: fake news packs a lot in title, uses simpler, repetitive content in text body, more similar to satire than real news," *arXiv preprint arXiv:1703.09398*, 2017.
[47] J. Fürnkranz, "A study using n-gram features for text categorization," *Austrian Research Institute for Artifical Intelligence*, vol. 3, pp. 1-10, 1998.
[48] A. Gupta, H. Lamba, P. Kumaraguru, and A. Joshi, "Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy," in *Proceedings of the 22nd international conference on World Wide Web*, 2013, pp. 729-736.
[49] F. Yang, Y. Liu, X. Yu, and M. Yang, "Automatic detection of rumor on Sina Weibo," in *Proceedings of the ACM SIGKDD Workshop on Mining Data Semantics*, 2012, p. 13.
[50] S. Tschiatschek, A. Singla, M. G. Rodriguez, A. Merchant, and A. Krause, "Fake News Detection in Social Networks via Crowd Signals," presented at the Companion Proceedings of the The Web Conference 2018, Lyon, France, 2018.
[51] S. Antoniadis, I. Litou, and V. Kalogeraki, "A model for identifying misinformation in online social networks," in *OTM Confederated International Conferences" On the Move to Meaningful Internet Systems"*, 2015, pp. 473-482.
[52] J. Kim, B. Tabibian, A. Oh, B. Sch, #246, lkopf, *et al.*, "Leveraging the Crowd to Detect and Reduce the Spread of Fake News and Misinformation," presented at the Proceedings of the Eleventh ACM International Conference on Web Search and Data Mining, Marina Del Rey, CA, USA, 2018.
[53] C. Shao, G. L. Ciampaglia, A. Flammini, and F. Menczer, "Hoaxy: A Platform for Tracking Online Misinformation," presented at the Proceedings of the 25th International Conference Companion on World Wide Web, Montréal, Québec, Canada, 2016.
[54] S. Jain, V. Sharma, and R. Kaushal, "Towards automated real-time detection of misinformation on Twitter," in *Advances in Computing, Communications and Informatics (ICACCI), 2016 International Conference on*, 2016, pp. 2015-2020.
[55] J. Zhao, N. Cao, Z. Wen, Y. Song, Y.-R. Lin, and C. Collins, "# FluxFlow: Visual analysis of anomalous information spreading on social media," *IEEE transactions on visualization and computer graphics*, vol. 20, pp. 1773-1782, 2014.
[56] K. K. Kumar and G. Geethakumari, "Detecting misinformation in online social networks using cognitive psychology," *Human-centric Computing and Information Sciences*, vol. 4, p. 14, 2014.
[57] S. M. Jang, T. Geng, J.-Y. Queenie Li, R. Xia, C.-T. Huang, H. Kim, *et al.*, "A computational approach for examining the roots and spreading patterns of fake news: Evolution tree analysis," *Computers in Human Behavior*, vol. 84, pp. 103-113, 2018/07/01/ 2018.
[58] C. Castillo, M. Mendoza, and B. Poblete, "Information credibility on twitter," in *Proceedings of the 20th international conference on World wide web*, 2011, pp. 675-684.
[59] Z. Jin, J. Cao, Y. Zhang, and J. Luo, "News Verification by Exploiting Conflicting Social Viewpoints in Microblogs," in *AAAI*, 2016, pp. 2972-2978.
[60] J. Ma, W. Gao, Z. Wei, Y. Lu, and K.-F. Wong, "Detect rumors using time series of social context information on microblogging websites," in *Proceedings of the 24th ACM International on Conference on Information and Knowledge Management*, 2015, pp. 1751-1754.

[61] J. Ma, W. Gao, P. Mitra, S. Kwon, B. J. Jansen, K.-F. Wong, *et al.*, "Detecting Rumors from Microblogs with Recurrent Neural Networks," in *IJCAI*, 2016, pp. 3818-3824.

[62] S. Kwon, M. Cha, K. Jung, W. Chen, and Y. Wang, "Prominent features of rumor propagation in online social media," in *Data Mining (ICDM), 2013 IEEE 13th International Conference on*, 2013, pp. 1103-1108.

[63] E. Tacchini, G. Ballarin, M. L. Della Vedova, S. Moret, and L. de Alfaro, "Some like it hoax: Automated fake news detection in social networks," *arXiv preprint arXiv:1704.07506,* 2017.

[64] W. Y. Wang, "" Liar, Liar Pants on Fire": A New Benchmark Dataset for Fake News Detection," in *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, 2017, pp. 422-426.

[65] W. Y. Wang, "" liar, liar pants on fire": A new benchmark dataset for fake news detection," *arXiv preprint arXiv:1705.00648,* 2017.