Fake News Detection Using Deep Learning

Engr. Rafi Ullah Khan

01-242182-008

Supervised by:

Principal Supervisor: **Prof. Dr. Shehzad Khalid**

Co-Supervisor: **Dr. Taimoor Khan**

Department of Computer Engineering

Bahria University Islamabad

2021

THESIS COMPLETION CERTIFICATE

Student's Name: RAFI ULLAH KHAN, Registration No. 01-242182-008 Program of Study: MS-CE Thesis Title: FAKE NEWS DETECTION USING DEEP LEARNING.

It is to certify that the above student's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for Evaluation. I have also conducted a plagiarism test of this thesis using HEC prescribed software and found a similarity index to 16% that is within the permissible limit set by the HEC for the MS/MPhil degree thesis. I have also found the thesis in a format recognized by the BU for the MS/MPhil thesis.

Principal Supervisor's Signature:

Name: Prof. Dr. Shehzad Khalid

Date: 08/10/2021

APPROVAL FOR EXAMINATION

Scholar's Name: Engr. Rafi Ullah Khan, Registration No. 01-242182-008, Program of Study: MS(CE), Thesis Title: Fake News Detection Using Deep Learning.

It is to certify that the above thesis has been completed to my satisfaction and my belief. Its standard is appropriate for exam submission. I have also conducted a plagiarism test of this thesis using HEC prescribed software and found a similarity index of about 16% that is within the permissible limit set by the HEC for the MS degree thesis. I have also found the thesis in a format recognized by the BU for the MS thesis.

Principal Supervisor's Signature:

Name: Prof. Dr. Shehzad Khalid

Date: 08/10/2021

AUTHOR’S DECLARATION

I, Rafi Ullah Khan solemnly states that my MS thesis Fake News Detection Using Deep Learning is my effort. It has no such material which is earlier published. All the mentions and necessary help in this study have been recognized. I affirm that the material in this research has not been used by me from Bahria University for another degree at any other institution.

Author’s Signature

Name of the Scholar: **Engr. Rafi Ullah Khan**

Date: **Oct 08, 2021**

PLAGIARISM UNDERTAKING

I, Rafi Ullah Khan solemnly declare that research work presented in the thesis titled "Fake News Detection Using Deep Learning" is solemnly my research work with no significant contribution from any other person. Small contribution helps wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero-tolerance policy of the HEC and Bahria University towards plagiarism. Therefore, I as an author of the above-titled thesis declare that no portion of my thesis has been plagiarized and any material used as the reference is properly referred /cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled even after awarding of MS degree. The university reserves the right to withdraw /revoke my MS degree and HEC and the University has the right to publish my name on the HEC/University website on which names of scholars are placed who submitted plagiarized thesis.

Author's Signature

Name of the Scholar

Engr. Rafi Ullah Khan

DEDICATION

To My Father, Mother, all my Family, and Teachers.

ACKNOWLEDGMENT

I am eternally grateful to Allah Almighty; the creator of all universes and everything that has ever existed. It would certainly have been impossible to complete my thesis without His guidance. I could feel the divine help all along with my thesis. Indeed, it sounded an uphill task, but Allah was always there whenever I found myself in any blind alley.

I want to extend my heartiest love to my parents, and family as well as a deep respect to my colleagues, who always provided me with moral support and always encouraged me whenever I got stuck at any point. I want to thank my supervisor, **Prof. Dr Shehzad Khalid** and my co-supervisor **Dr Taimoor Khan** for their immense guidance that played an instrumental role in shaping up my dissertation work. Dr Taimoor taught me how to carry out research effectively and veritably, he is the architect of this work and I want to give him maximum credit for that.

Lastly, I want to pay my earnest gratitude to my family, friends and colleagues who always boosted my morale, especially my brother **Dr Sami Ullah Khan** and they are the true heroes and equally deserve all the recognition and praise. There are certainly unsung heroes as well and I can't name them all, but I do like to take this opportunity to send my warmest regards to whoever played his part in making me what I'm today. I owe it to all of them.

Abstract

Classification of fake news content is one of the challenging problems of Natural Language Processing (NLP). In classification of fake news detection, the classes of true and fake news are predefined in which the news are assigned based on model's judgement. The increase in social media gave an edge to spreading of fake news easily. it has now become one of the considerable menaces to journalism, democracy and freedom of expression. In this era, fake news has emerged as a world topic, and it has become of major concern for the people to know the authenticity of a news content over the social media. The existing content-based approaches such as rule based, probabilistic and machine learning are used for classification. These approaches are far from achieving acceptable accuracy with fake news detection due to the complex nature of the news content that is generated to mislead the audience. These models require handcrafted features, which has the possibility of missing out the important features or considering the unimportant features. Secondly, these traditional models lack the ability of memory element to keep the track of previous words as well as current appearing words also known as words dependency, which is one of great importance in the classification of fake news. In this research, we have proposed classification of fake news model that comprise of news content representation scheme also known word embeddings and deep learning model that represent the news articles as latent features of the text. The proposed model for classification of fake news is a blend of two DL models consisting of 1D Convolution Neural Network (CNN) as feature extractor and LSTM as classification model. Our model outperforms the state of art by a well-off accuracy on three known datasets of fake news detection such as 98.6% on FakeNewsNet datset, 97.30% on ISOT dataset and 90.08% on FA-KES dataset.

Keywords: Classification, Fake News Detection, Natural Language Processing, Deep Learning.

Table of Contents

Contents

List of Figures

List of Tables

List of Symbols and Abbreviations

NLP	Natural Language Processing
1d CNN	1 Dimensional Convolutional Neural Network
LSTM	Long Short-Term Memory
TF-IDF	Term Frequency-Inverse Document Frequency
PCFG	Probabilistic Context-Free Grammar
SVM	Support Vector Machine
LSVM	Linear Support Vector Machine
KNN	K-Nearest Neighbor
DT	Decision Tree
KG	Knowledge Graph
SGD	Stochastic Gradient Descent
LR	Logistic regression
LDA	Latent Dirichlet Allocation
RST	Rhetorical Structure Theory
LIWC	Linguistic Inquiry and Word Count
CBoW	Continuous Bag of Word
ANN	Artificial Neural Network
BRNN	Bidirectional Recurrent Neural Network
BLSTM	Bidirectional Long Short-Term Memory
DL	Deep Learning
GRU	Gated Recurrent Unit
GloVe	Global Vectors for Word Representation
IDF	Inverse Document Frequency
RNN	Recurrent Neural Network
CSI	Capture, Score, Integrate
NLTK	Natural Language Tool Kit

CHAPTER 1

INTRODUCTION

Fake news isn't the conspiracy of new time, it has been from long since people started to gain political and social fame. With the increase of social media, it is now one of the greatest threats to journalism, democracy, and free speech, due to which it has emerged as a world topic and has attracted the public towards its popularity at this time [1]. Before the social media popularity, it was not easy to create and publish fake news using conventional media such as television and newspapers but now it is way easier and cheaper to creating and spreading fake news due to the rise and availability of social media. News quality over social media is less than traditional media due to the mentioned advantages of social media. Fake news dissemination becomes easy and the physical barrier among individuals is overcome by the ideal platform of social media to accelerate this process by voting, sharing, forwarding and reviewing, the users can easily discuss news online [2]. Psychological and social factors acted as one of the main roles in getting trust of the public and in spreading fake news. With a high amount of deceptive information, it becomes hard for humans to differentiate between truth and falsehood of a news. Communication, social and psychology studies show that the ability of humans in deception analyses and detection is not more than a chance with the accuracy of 55% to 58% in range [3]. Fake news in large volume, i.e., news or articles written and produced over social media for the purpose of financial and political gain. To mislead the reader, fake news is written intentionally. This makes it difficult to detect and classify because it is diverse in terms of media platforms, styles, and topics. Non-factual claims within the incorrect context are supported by citation of true evidence for spread of fake news [4].

Fake news is diverse, and its definition is not universal, even in journalism. There is a need of definition for fake news that is accurate and clear to help in compact infrastructure for analysis and evaluation study. Several concepts are out there that somehow fall in the fake news category. These terms are differentiated based on three main characteristics. Firstly, authenticity that shows if the news is false or not. Secondly, intent that describes the intention of news, i.e., whether it was created with bad intention or not. Thirdly, the given information is news or not. Some of the studies connect these terms with fake news shown in figure 1, such as maliciously false news that has false authenticity created with bad intention to mislead the public for evil benefit of specific people [1] [5] [6]. Clickbait news to increase the curiosity of consumer, one needs to click the link which leads to full article or to incite confusion [7]. False news is created with false authenticity and intention (bad or not) to mislead the consumers [8]. Satire news are the news with false authenticity, but not bad intentions created for humorous purpose [9]. Disinformation also called deception is information that could be news or not with bad intention and false authenticity that aims to mislead the consumers [10]. Misinformation is false information with intentions (bad or not) to mislead public [11]. Rumor is the type of information created with uncertainty and is doubtful [12].

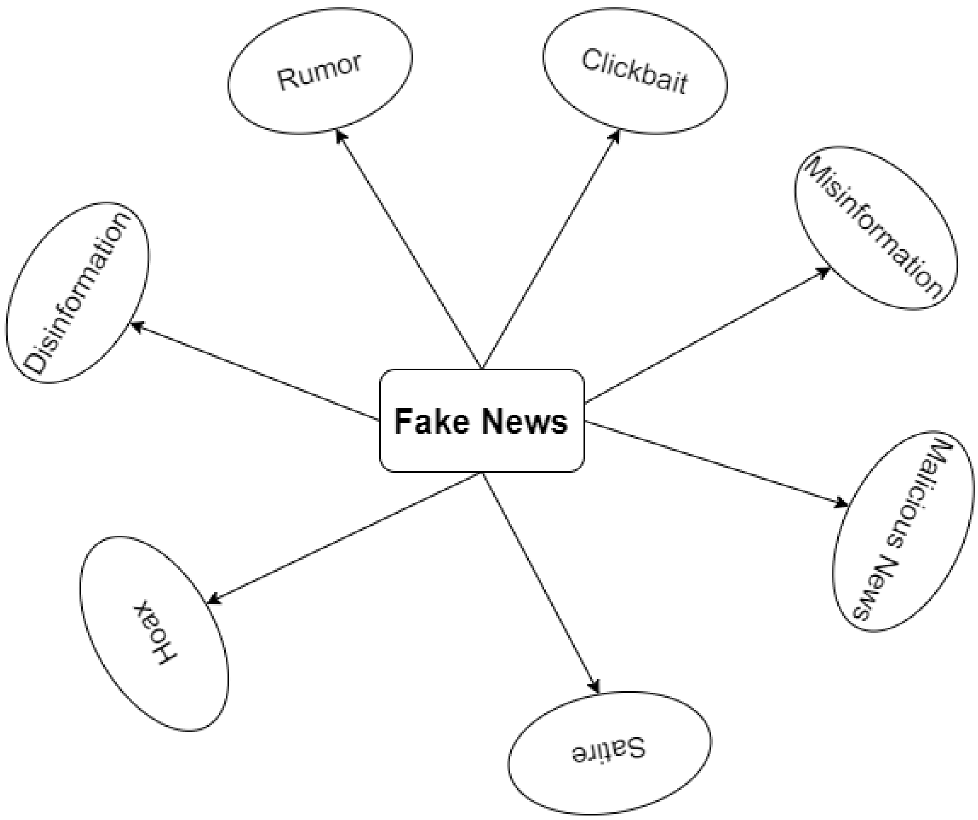

Figure 1.1 Types of news associated with Fake news

Detection of fake news is divided into three categories shown in figure 1.2, i.e., news content based, social context-based and propagation-based detection [2][5]. In news content based, fake news is differentiated with the help of clues that news contents contain. Features in news content-based detection are extracted as linguistic and visual based. In linguistic features, writing style and headlines of the news is captured that frequently appears in content of fake news, such as syntactic and lexical features [13]. Fake images that are created to mislead users are identified with the help of visual-based features that nabbing some characteristics of fake news images [14]. Fake news is approached by knowledge based, style-based and latent features perspectives. Knowledge based detection aims to check the veracity of a claim in news content by fact checking with use of external sources. Fact checking process assigns the claim a truth value for veracity in a particular context [15]. Style-based detection aims to capture the manipulators for detection of fake news in news content writing style.

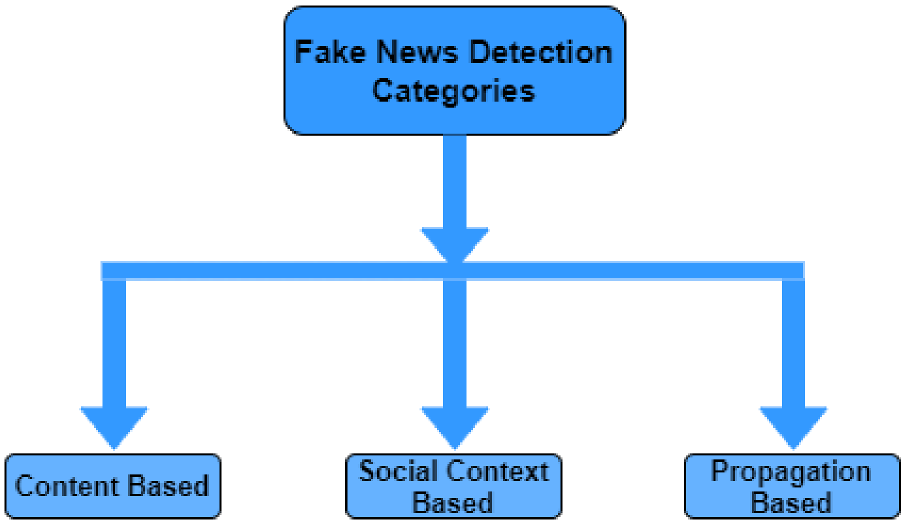

Figure 1.2 Fake News Categories

Detection of fake news by using the social context strategy require enough social context information which is gathered from social media news ecosystem. it comprises of three core entities, user profiles, news pieces and publishers. News pieces encloses enormous information that helps in detection of fake news and there are social groups that cause echo chamber effect (people having similar interests and instincts). Features for social context are post, user and network based. Measurement of credibility and characteristics of users includes features that are extracted from user profiles [16] [17] [18] [19]. In user based, features from user profiles that represented the naïve user, which believe in fake news and experienced users that can sense the fakeness of a news, were selected to differentiate fake news [18]. Social media engagements, credibility of users and truth of news were used to determine the opinions of users toward the authenticity of a generated news in an unsupervised manner [19]. In post based, features include the responses of users on social media in terms of stance i.e., towards or against some news [24] and topics [20]. In network based, specific networks are constructed to extract features, such as stance network that is based on the similarity of stances of the users, co-occurrence network built on basis of engagement of users toward a specific news, and diffusion network that tracks down the start of news spreading [21]. Social context-based modeling includes stance-based detection that aims to figure out the original news article veracity by utilizing the post content of the user [22].

Propagation based fake news use the features that are extracted from the information of dissemination of fake news, i.e., the way fake news is propagated and the users that spread it. Faker news based on propagation methods lead to interrelations of related social media posts, thereby predicting the credibility of news. The elemental assumption is that news events credibility is highly correlated with related social media posts credibility. Propagation approach is narrowed down to cascade and network detection. Cascade based fake news represents a certain fake news propagation in a tree structure spread by users on social media. This tree structure has two parts, one is the root node which is also called the creator or initiator that firstly posts the fake news on social media and the other part includes the child nodes of the creator (Parent node) that forward/post it after it was posted by root node. Fake news can be distinguished with utilization of cascade similarity, in which similarity of a cascade is computed to that true or fake news and differentiating true news from fake news using information representation by properly representing its cascade. In network approach, propagation of fake news is captured indirectly with the help of flexible networks [26]. Homogeneous, heterogeneous and hierarchical flexible networks can be established for the dissemination process. In homogeneous networks, a single type of entity network is built, e.g., post or event [23] while in Heterogeneous networks, credibility of different entities types are involved, such as events, sub-events, and posts [24]. In hierarchical networks, nodes and edges form relational set-subset hierarchy [25].

Most of the fake news datasets available are comprised of news content only but researchers are working for the dataset that has content, user context, social engagements in a network form. Some datasets have metadata information of fake news, and some has text and images mix information about the fake news. As the detection of fake news with the social context information and its propagation patterns are difficult to carry out with the limiting constrains, we are working on the content-based approach to enhance the accuracy of fake news detection.

Many techniques are evolved with the passage of time and some new ones are suggested to enhance the process. Over the past Machine learning and neural network approaches are applied in multiple researches to solve this inherent problem but due to not getting an acceptable accuracy the researches are still working on this problem to find a system that can solve the fake news detection issue with an acceptable accuracy. The existing approaches of artificial intelligence and traditional machine learning are far from achieving acceptable accuracy with fake news detection due to the complex nature of the problem, diversity of domains that may be targeted for fake news. A major challenge for automated fake news detection is the features extracted for detection that's helps in achieving a well-off accuracy.

Our motivation of research in this area leads us to explore the different schemes and models that leads us to propose a system that surpasses the existing approaches to solve the problem of fake news detection. Dealing with the task of fake news detection requires deep insight of the topic, subject and NLP. Detection of news as fake or true based on content approach is a hard task as the news could be a humorous statement. However, deep learning models, word embeddings, and other models such as sequence models are used to develop models that involve advancement in NLP [26]. In this research, we have proposed an approach with deep sequence model using embedded vectors for fake news classification. Our proposed model achieved a plausible accuracy for predicting fake news from true news on three of the fake news datasets published differently. First one is FakeNewsNet that is comprised of two datasets collected from PolitiFact and BuzzFeed [27]. The second dataset we is ISOT that is gathered from Retuers and Kaggle based on real world sources [28]. Third dataset is FA-KES, that was gathered and created around the Syrian war that has the true and fake news of the war from multiple news outlets [29].

The classification accuracy, precision, recall, and F1-score are the evaluation metrics targeted in our experiments. Our proposed framework performed well in comparison to the existing models on three known datasets as mentioned and explained above. On BuzzFeed dataset (FakeNewsNet), our proposed model achieves improvement of 6.3%, Similarly by on PolitiFact dataset (FakeNewsNet), our proposed framework outperforms the TriFN model by an improvement of 10.8% in term of accuracy. An improvement of 5.3% and 31.99% accuracy is shown by our proposed approach on ISOT and FA-KES datasets, respectively. Our proposed framework outperformed the state of art models by overcoming word dependency and feature representation.

1.1 Problem Statement

This research aims to solve a very inherent problem of content based fake news classification. The existing content-based approaches are far from achieving acceptable accuracy with fake news detection due to the complex nature of the news content that is generated to mislead the audience. These models focus on handcrafted features and machine learning or artificial intelligence methods which also lack the ability of words dependency, which are of great importance in the classification of fake news.

1.2 Motivation

Our motivation for working on the problem of fake news detection comes from the ongoing social media usage for spreading fake news. During the times of elections and beside the elections fake news are spread over the social media to change the ideology of audience towards a political party or to benefit some organization. As the reach of social media is easy for every individual in today's world, it has become too easy to spread fake news to benefit or harm someone.

1.3 Research Objective

Research objectives of our thesis is to classify the fake news from true news and to develop a system that detects fake news from the clues of the news. This system will not only target the politics news but also the news from other domains such as media, sports etc. Developing a system that detects the news on an early stage of its spread will help to restrain the audience from further spreading it or the naïve audience to be deceived.

1.4 Research Contribution

Our contribution to the previous work in fake news domain is that we have focused on two aspects of detection of fake news, first is to extract the features from the news content as of latent features representation by extracting the features using deep learning model instead of creating the feature map using handcrafted features that require time and domain knowledge. Secondly, word dependency is one of the main factors in detection of fake news which lacks in traditional machine learning models. We trained our model to learn the word dependency from the news. Our proposed model achieved a plausible accuracy.

1.5 Organization of the Thesis

In chapter 2, we undergo a study of fake news related works. We have briefly discussed literature and recent development which help us to sort out the problem and its solution for the classification of fake news. The chapter 3 presents the complete methodology, details of the different enhancements, and major phases in developing a fake news detection model. In chapter 4, we have presented experiments and results that are included with exploratory data analysis, proposed model training, and it's testing with the help of graphs, and tabular results. Finally, in chapter 5, we have concluded the thesis and explored potential enhancements and developments of the existing systems as future works.

CHAPTER 2

LITERATURE REVIEW

False information and false news have been extensively used in computer science research however it has its roots from social sciences i.e., “fake news”. Fake news is defined as crafted information that resembles news media content in appearance but differs in organizational procedure and goal. In turn, fake news outlets lack the news media's editorial norms and mechanisms for assuring information's authenticity and reliability. The Internet has most principal enabler and primary conduit of fake news [1]. Different studies have been conducted to find trends in dispersal of disinformation on social media. According to one such study, through the end of 2016, user interactions with fake news risen substantially on both Twitter and Facebook. [30]. The fake news repositories are publicly provided to promote research in this direction [27]. Fake news detection is a very challenging job which requires consulting a variety of disciplines and research areas. Overview of the research taken out in the field or domains of fake news detection is classified according to its type i.e., content based, social context based, and propagation based as shown in Fig 2.1.

2.1 Content Based Detection

The initial focus on false information detection was carried out on subjective content [31]. Different approaches were considered for this purpose to explore user and review based features that would discriminate between false and genuine information [32]. Some of these features were also employed for detection of fake news as well. In [28], authors introduced ISOT dataset. The news in dataset are based on entirely real-world sources. On the ISOT dataset, the author used four n-gram and ML models for detection, including the K-Nearest Neighbor (KNN), Support Vector Machine (SVM), Linear Support Vector Machine (LSVM), Stochastic Gradient Descent (SGD), Decision Tree (DT) and Logistic Regression (LR). TF-IDF and LSVM as the feature extractor and the classifier respectively achieved best accuracy i.e., 92%. In [33], Term Frequency-Inverse Document Frequency (TF-IDF) of bigrams and probabilistic context-free grammar (PCFG) detection were employed in conjunction with various models such as gradient boosting and stochastic gradient descent.

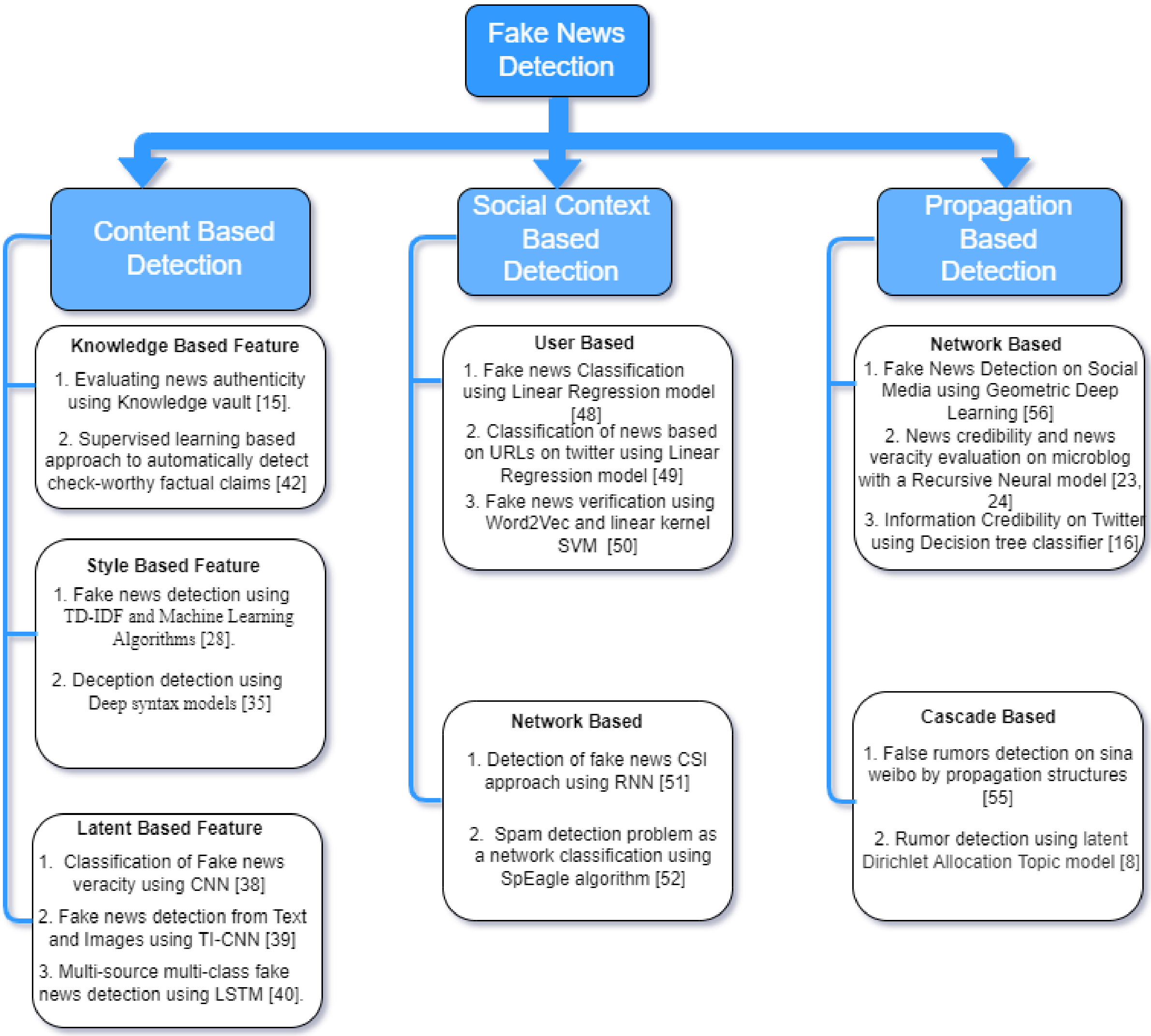

Figure 2.1 Fake news detection domains

The TF-IDF of bigrams using the SGD model classified fake news achieving 77.2% in term of accuracy. However, because these are distinct to the training dataset, utilizing only the vector-based approach cannot assess individual features and train classifiers. In [34], the authors applied multiple models of machine learning and neural network including KNN, Decision Tree, Logistic Regression, Bernoulli Naïve Bayes, Multinomial Naive Bayes, SVM, LSVM, Perceptron, Neural Networks on the dataset FA-KES which was introduced around the news of Syrian war. Though the models did not perform well on the dataset and the best accuracy achieved by the multinomial naïve bayes models which was 58.09%. Natural language processing (NLP) models such as Rhetorical structures and Deep syntax are applied to detect fake news.

Rhetorical Structure theory (RST) plays its role where one wants to differentiate truthful sentences from deceptive ones. RST is implemented using vector space model (VSM) for the detection of fake news [35].

In one of other studies a method used for feature extraction for fake news detection is LIWC which is extensively used for extraction of sentiment lexicon that fall into psycholinguistic types. It is built on a broad vocabulary of words that learns a feature vector based on psychology and deceit [36]. RST and LIWC both techniques were employed by the author for fake news detection [37]. Multiple machine learning algorithms were applied on the features extracted by RST and LIWC methods on FakeNewsNet dataset. The algorithms applied are Logistic Regression (LR), Random Forest (RF), Decision Tree (DT), Naïve Bayes (NB), XGBoost, AdaBoost, and Gradient Boosting (GB). Best accuracy achieved on BuzzFeed dataset for RST and LIWC was achieved by AdaBoost algorithm i.e., 63.3% and Random Forest i.e., 70.9% respectively.

As the deep learning models performed very well in multiple disciplines and proved the worthiness of its models, in one of the studies, a deep neural model is found in fake news detection in which CNN is used for classification of fake news veracity on a dataset named liar containing short claims that are annotated by the editors and journalists [38]. In [39], the author trained a convolutional neural network model TI-CNN with images and text information. Lexical divergence and sentiment were used for the text and for images, observance was made that face news images had much irrelevancy in images and true news had more images having faces. For latent features extraction from visual and textual data two parallel CNNs were used in the model achieving precision rate of 0.92 and recall rate of 0.9227. A large number of examples in the dataset are required for CNNs and dealing with images and texts both is computationally expensive.

Fake news detection model also represented for Multi-class Multi-source, in which local patterns in a claims text is analyzed by CNN and the entire text document is analyzed by LSTM for temporal dependencies as for long sentences LSTM works better [40]. Content based detection work extended by detection of satirical news using SVM based algorithm, five source features were selected, the predictive features including Humor, Grammar, Punctuation, Absurdity and negative affect, which was tested on 360 news article datasets. The approach of their detection started with classification of topics followed by classification of sentiment. Features selected based on humor, heuristic, and absurdity. Extracted features given to SVM linear kernel model to predict the satire news from the articles. The best combination of their predictive feature came out as grammar, absurdity, and punctuation, which yield a fine precision and F1 score [41].

Content based fake news detection by knowledge extraction is taken in this field. Knowledge based detection approach of news content attempts to directly assess authenticity of the news content by correlating knowledge extracted from news data that is to be verified stored in a Knowledge Graph (KG) like Knowledge Vault [15]. Knowledge based model approach used to fact check the worthy claims from sentences and specially from presidential debates transcripts. The problem was modeled as the classification in a supervised learning framework. Labelled dataset created from presidential debates, the labels used as Important factual claim, not factual claim and unimportant factual claim. Different classification models were trained and tested over the labelled dataset such as SVM, Random Forest classifier (RFC) and Naïve Bayes classifier (NBC). To avoid the overfitting of the models, features were selected using features selection. The features such as Sentiment, Length (of sentences) and Part of speech were selected for the models. Accuracy of the models showed promising results after feature selection [42].

The style based fake news detection approach was used to detect the hyper-partisan news that is defined by the news which is extremely one sided. They analyzed that the hyper-partisan news is distinguished from mainstream news on the basis of its writing style. They used Unmasking as a state of art technique to find the originality document to its author and if it is one sided or not. The features selection was based on stop words, part-of-speech, and word frequency [13]. Only using linguistic aspects to detect fake news from articles is used in many studies without using fact checking. Fact checking is to be known for misclassification of fact temporal fake news. combination of these two i.e., linguistic and fact checking characteristics can perform. As the news on social media increases and needs to be tackled. Crowdsourced based KG was proposed to fact check the news events timely [43].

2.2 Social Context Based Detection

Many of the researchers used the content of the news and applied different techniques to detect fake news detection. However, there are some vital differences between the content and context, which necessitated the need of different approaches towards fake news detection [44]. In contrast to subjective fake content, fake news is based on factual information. What makes fake news detection a lot more challenging is its growing impact as it propagates on social networks. Different techniques proposed and experimented to classify or identify real news from fake news including social network analysis and data mining methods. The authors presented TriFN models for classification of false news which yields an accuracy of 86.4% on BuzzFeed and 87.8% on Politifact fake news dataset [45].

To analyze input text characteristics and to find the idea this structure can work by defining the functional relation such as purpose, evidence and circumstances between text units. News text is converted to vectors using a vector space model then compared to center fake and true news in higher dimensions of RST space, where dimensions in vector space shows the relations in text. In deep syntax, sentences are converted or transformed to some rules so that syntax structures could be described by implementing deep syntax models with probabilistic context free grammars (PCFG). Deception detection rules can be developed such as grandparent and lexicalized/unlexicalized rules on the basis of PCFG [46]. In [47], Authors conducted a compelling attempt to employ photographs for news verification by utilizing visual elements such as coherence and clarity score, as well as statistical features of images such as image ratio and count. Using these features of the images, they attained 83.6% of the verification accuracy. When compared to approaches that simply employ non-image features, accuracy of this approach was boosted by around 7%.

Users' interaction on social media is an important subject to be studied for fake news detection. As negative users always share or post hoaxes (news based on falsehood) and interact with likeminded people. Similar work taken out to identify hoaxes on the user interaction basis on Facebook. The problem is treated as a two-way classification problem. Where user interaction with posts or features is considered as a linear regression problem and second one is crowdsourcing technique in which training data set is available, but users are not predefined [48]. The proposed approach was extended by [49], where authors used the URLs of the tweets classified fake news from real news using a logistic regression model. In [4], the authors created a technique for detection of fake news by combining structural, user, content-based and temporal data in popular Twitter threads. Polarity, subjectivity, and disagreement were chosen as content-based features. Their technique’s scope is confined to Twitters’ extensively re-tweeted dialogues threads, and high number of tweets in real life are re-tweeted. Their model performed with an accuracy of 65.29% when applied to the BuzzFeed dataset. A system of stance detection was implemented which included sentiment analysis as a baseline for its detection. To classify a stance, it is a much more important task to determine the sentiment of a tweet, alone stance is not sufficient for classification problem of tweet. They built a linear kernel SVM dependent on features extracted using word embedding technique that is Word2Vec Skip-gram model from unlabeled data and from training instances i.e., character and word n-grams [50].

A hybrid deep model was implemented named CSI (Capture, Score, Integrate) that was composed of three modules. RNN was used with doc2vec feature module for the first module to determine temporal patterns and to capture these patterns of user's activities on texts. Second module worked based on the behavior of the user's graph to learn characteristics of the source implemented using a neural network. Their work includes text, interaction, and source user data. This method predicts fake news with an accuracy of 89.2% in the Twitter dataset and 95.3% in the Weibo dataset [51]. These modules combined with a third module that is classification of articles as fake or true. The problem of fake news detection is also attempted with graph-based techniques. The idea is to represent the users and their news subjects as nodes and link them with edges bearing the news. Following a belief propagation approach, bad users would discuss subjects that would in turn attract more fake news and vice versa [52]. Thus, propagating from few users as fake news promoters or subjects as fake news targets, the propagation mechanism can help in identifying more such users through their associations [53].

2.3 Propagation Based Detection

Social content and social context can be of a good combination but considering the fast-spreading nature of fake news on social media requires the consideration of the dynamic nature of online social media networks. Therefore, it is equally important to observe structural changes in a propagating network and to analyze propagating paths of news with respect to time for the flow of fake news. Moreover, different attempts are made to model fake news and to model their propagation so that such attempts may be identified in future [54]. News cascade is another way to directly represent news propagation in which post-repost relationships are presented by tree structures on social media for news articles like twitter's tweet-retweet relationship [26]. Based on news cascades, in [8], authors studied the asymmetrical diffusion of fake and true news articles on Twitter distributed from 2006-17. The authors calculated the information distance between the rumor tweets and all preceding tweets that users were exposed to before retweeting the rumor tweets using a Latent Dirichlet Allocation Topic model (LDA) trained on 10 million English-language tweets.

News cascade-based research is extended with user (naïve or malicious users) role, stance (agree or against) and expression of sentiments in user posts. The assumptions were made that fake news structure differs from true news structure for which random walk graph kernel was developed so that news cascades similarity could be measured, and fake news is detected on measured similarity [55]. In another study the problem was served in supervised learning by assessing news credibility with the help of feature selection from tweets, user profiles and news cascades [16]. A stance graph-based model approach was used for fake news detection from user posts by correlation of stance in a graph optimization groundwork in which relationships between user (spreaders) and user posts, publishers and news articles were extracted [23][24]. Recursive neural models such as top down and bottom-up tree structured neural networks were proposed to learn the representation and classification of rumors, which comply with tweets propagation layout. The problem was served in a supervised learning framework in which a Geometric deep learning approach was used on graph structured data to learn the fake news propagation pattern [56].

Fake news detected is also conducted for theory driven models, news content investigated at different levels such as lexicon, syntax, semantic and discourse levels. At each level news rely on theories that are well-established in forensic and social psychology [57]. If the network information and attributes of source user are provided, such information can be of worth in making a more accurate representation of the news′s credibility. However, due to privacy issues of users, it is not always feasible to get most of this meta data regarding users and their connections. Despite the fact that fake news writers aim to frame the information in such a way that it sounds authentic, false news can be spotted by examining some general textual features.

Table 2: Related works table of Content Based Fake News Detection

Author & Year	***Methodology & Dataset***	***Accuracy***
Ahmed et. al. (2017)	*N-gram with Machine Learning classifiers (ISOT Dataset)*	*0.92*
Wang et al. (2017)	*Support Vector Machine, CNNs (LIAR Dataset)*	*0.255 0.270*
Elhadad et. al. (2019)	*N-gram with Machine Learning Classifiers (MNB) (FA-KES Dataset)*	*58.09*
Perez-Rosas et al. (2017)	*Linear SVM Classifier (Politifact, BuzzFeed Dataset)*	*0.811 0.755*
Karimi et. al. (2018)	*Multi Source Multi Class-CNN and LSTM (LIAR + PolitiFact Dataset)*	
Shu et al. (2019)	*TriFN (Politifact, BuzzFeed Dataset)*	*0.864 0.878*
Yang et. al. (2018)	*TI-CNN (2 parallel CNNs) (Images and Text Dataset)*	*0.92*
Hassan et. al (2015)	*Knowledge Based Model – SVM, RFC, NBC (PolitiFact & Channel4 Dataset)*	*0.737*

2.4 Critical Analysis of Research in Fake News Detection

Over the past, Multiple methods and techniques were used for fake news detection but researchers found a need to attain a sound accuracy in fake news classification process. Due to the complex nature of fake news attaining a well-off accuracy is always the challenging task in fake news detection research. If we consider Knowledge graphs (KGs), When depending on the knowledge included within a news content to identify if it is fake or not, the knowledge obtained from it can be compared to the one in KG as ground truth's source. The formation of such a graph structure is still a work in progress, notably for the detection of fake news. First, according to news definition "newly received or significant information, particularly regarding recent incidents," implying that time should be taken into account while creating such knowledge graphs [26]. Second, these KGs are quite often incomplete, necessitating the development of methodologies for knowledge implication, and fake news defined as “news that is intentionally and verifiably false” [5]; such knowledge-based techniques can assist in verifying news truthfulness but cannot verify the intentions of those who create news articles [26].

Current style-based techniques of news content approach focus on capturing the non-latent features from the content of the news, such as distribution of n-grams [28], TD-IDF based word level statistics [33] and LIWC features usage [36]. The limitations of using style-based techniques for extraction of features has some of the limitations e.g., similarity of a document is computed directly by TF-IDF in the word-count space, for extended vocabularies it may work slower. Secondly, it is assumed that different words count give independent proof of similarity. Thirdly, utilization of semantic similarities lacks between words which is most important for the detection of fake news to consider the relationship between similar words. In n-gram models the long-range dependencies are not captured and are dependent on having a corpus of data to train from i.e., sparse data for low frequency affect tags adversely affects the quality of the n-gram model. LIWC is used to measure the bag of words sentiment.

Approaches based on ML models have captured lots of recognition in recent years. By observing the data these approaches learn and classify tasks that are text based. A pre-labeled data is used as examples for the training of ML algorithm that learns the patterns/association between text and their labels. Most of traditional ML models works on the procedure of two steps which include feature extraction and prediction. In first step, dataset is used to extract the required features these features are extracted by handcrafted method then these handcrafted features are then given to the classifier models for prediction of class to which the unseen data/information belongs to. Bag of words (BoW) is one of the favored handcrafted features with the inclusion of expansions. ML algorithms such as SVM, Naïve Bayes, HMM, random forests and gradient boosting trees are used for classification. There are some limitations of two step approach. For example, depending on the features extracted by handcrafted method needs monotonous feature engineering and to obtain well off performance also requires analysis. Moreover, the robust reliance on domain expertise for features designing turns models to complexity when it comes to generalize to new problem. Due to the predefined feature set, we cannot take full advantage from training dataset of large amount.

Models that are based on deep architectures also called deep learning models have shown a remarkable popularity in achieving plausible accuracies in numerous tasks of text classification. These tasks include news categorization, natural language inference, question answering and sentiment analysis. In the case of sequential data, it is more evident that deep learning models performed better than machine learning models for many problems. Classification of text is one of the matured problems of NLP, also known to be as categorization of text, the main objective of it is to assign correct labels to unit of texts i.e., sentences, paragraphs or could be a document. Text is considered as an immensely rich information source. This quality of text and because of unstructured character it becomes challenging and time consuming to extract the real meaning behind or insights from the text. The main distinction between traditional pattern recognition and deep learning approaches is that DL automatically learns features from data rather than using handcrafted/constructed features. Deep learning discovers dispersed feature representation of given data by combining lower-level characteristics to generate more abstract and higher level characteristics [58].

This trait of DL models plays important role in accuracy improvement, as by adopting handcrafted feature, the chances are more to not consider all important features or some of the features that plays important role are missed out. Another attribute of DL models such as LSTM (Variant of RNNs) is that these models work on word dependencies which is a must part of text classification task because to analyze different aspect of a news it is important to consider the words dependencies appearing in a sentence. As mentioned in [58], DL models starts from very low-level features to form higher level features in that way more features are considered for the problem. Sequential data is processed using RNNs. In classical NN models, the data is operated initially at the input layer and forwarded to hidden layer and then to output layer. The nodes of each layer are not connected but the layers are fully connected with each other. RNNs or variant of RNNs are generally preferred for tasks that need processing of sequential inputs/information, such as tasks related to voice and language [58].

2.5 Models Employed for Classification in FND

Artificial neural networks are not able to handle sequential or temporal data because they lack memory elements. They have fixed architecture, which means that they handle the inputs for the present time step only. They don’t keep the input at the previous time step. Secondly, it requires domain knowledge to extract the features set which is a hectic and unpredictable method because one can pick the features that are not of much worthy for the task leaving some of the important features. To solve this problem, we come up with a solution which is Deep learning techniques i.e., 1d Convolution and LSTMs, they are biologically realistic as we humans have persistence in our thoughts, and we have some memory element too in LSTMs. RNNs have memory storage elements that is a plus point when one is dealing with sequential data, but it comes with a lacking element of long-term dependencies i.e., called vanishing gradient problem for which we have selected LSTMs to deal with the long-term dependencies problem. We can't use feed-forward networks when the data has interdependency and some interconnection between them because they need some memory for previous, current, and next time steps so that is why we use a variety of RNNs models (LSTMs) that have been employed in literature to solve the problem of fake news detection. Some of the major models are included in the following list.

2.5.1 Convolution Neural Network (CNNs)

Convolutional Neural Networks originally were developed initially for computer vision problems, have been found to perform well on text classification tasks and other typical NLP tasks. First time CNN were used in sentiment classification task, in which author modeled the CNN architecture for text-based task and performed 1d convolution [59]. Convolution in NLP tasks, have a 1-D array when applied to text instead of images. The architecture of ConvNets in textual task replaced with 1D convolutional and pooling operations. Contemplate window of some words, then the convolution filter is applied to each window, resulting in scalar values r_i. The pooling technique is used to concatenate or combine the vectors produced by various convolution windows into a single 1-dimensional vector space. The notion that ConvNet is merely a feature extractor - most likely to a fully connected layer for classification. A detailed review of the CNN architecture is given in chapter 3 methodology.

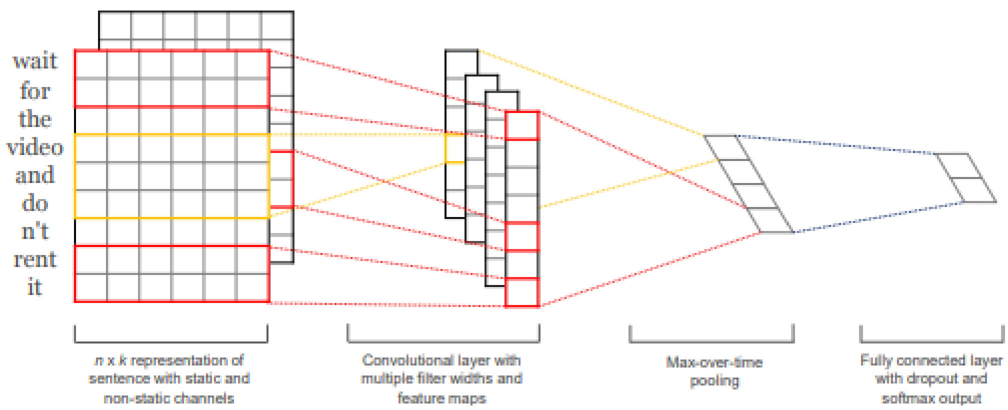

Figure 2.2: Model architecture with two channels for an example sentence. [59]

2.5.2 Recurrent Neural Network (RNNs)

Recurrent neural networks (RNNs) are basically a class of neural networks that permit previous outputs to be used as inputs while having multiple time steps and hidden states and produces an output. Usefulness of these models are worthy when dealing with sequential tasks such as text. Formally RNN takes an input, processes it through multiple time steps and hidden layers, and produces an output. In Fig.2.3 (a) that it is learning in a loop like structure. it is showing the activation of one layer fed back to the network and the next time step with a delay of one-time step. The RNN's hidden state that has functionally dependent on the hidden state at the preceding time step. As we know RNN can solve many-to-many, one-to-many, and many-to-one problem.

An example of many-to-many problems and multiple hidden layers is shown in Fig.2.3 (b). The functional dependencies for the hidden state are input, hidden state at the previous time step, same layer, hidden state at the current time step, and previous layer. The hidden state can have only two functional dependencies at maximum. The main functional dependencies of hidden layers are given below. whereas l represents the depth of hidden layer, h shows a symbol of hidden layers, l shows layer number in Fig.2.3 (b) the first hidden layers are l_1 and the second is l_2 , and t shows the time step. Current input x<t> is actually x_t and the f means it is functionally dependent in order to compute. Now for let $l = 1$ is 1st layer which is directly connected with the input layer, and it also depends on the hidden layer with previous time h_{t-1}^{l} see Fig. 2.3 (d). Now for $l > 1$ so here h_t^l depends on the previously hidden layer at the same time step h_t^{l-1} and activation of $t-1$ at same layer is h_{t-1}^{l}. This is actually a functional dependency of the hidden layer see Fig.2.3 (c).

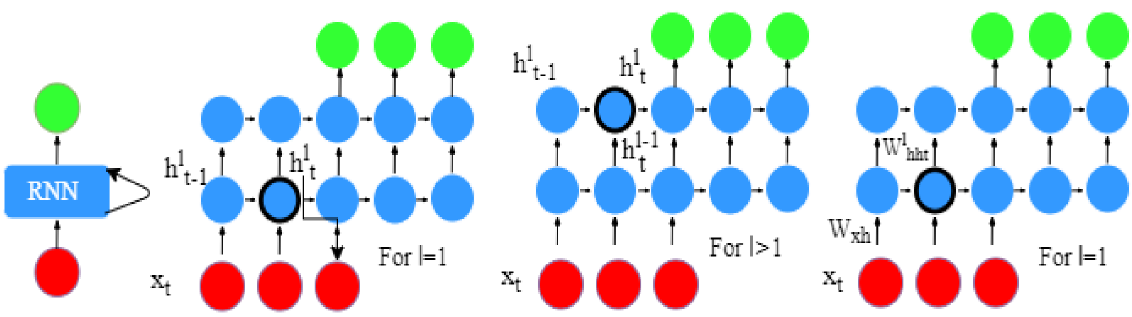

Fig 2.3 RNN Unfolding a, b, c, d [60]

$$h_t^l = f_x\left(h_{t-1}^l, x_t\right) \qquad for\ l = 1 \qquad (2.1)$$

$$h_t^l = f_x\left(h_{t-1}^l, h_t^{l-1}\right) \quad for\ l = 1 \tag{2.2}$$

Then functional dependencies are multiplied with weight matrices.

$For\ l = 1$

$$h_t^l = tanh\left(W_{hht}^l\, h_{t-1}^l + W_{xh}\, x_t + b_h^l\right) \tag{2.3}$$

$For\ l > 1$

$$h_t^l = tanh\left(W_{hht}^l\, h_{t-1}^l + W_{hhd}\, h_t^{l-1} + b_h^l\right) \tag{2.4}$$

$$y_t = g(W_{hy}\, h_t^l + b_y) \tag{2.5}$$

RNN – Common Activation Function: For the hidden layer, $tanh$ is used which compresses value between 0 and 1. For the output layers, *Softmax* or *Sigmoid* is used which compresses the value between 1 and -1.

Vanishing (Exploding) Gradient occurs when the algorithm back propagates as shown in Fig.2.4 (a) the value of gradient will be very small sometime may become large that is called Exploding Gradient. In case of RNNs, we have multiple time steps and error is transferred through all these time steps as illustrated in Fig.2.4 (b). So, once it propagates the error and apply the chain rule, initial layers are not able to learn. This is a very common problem in RNNs and the solution to this problem is Long Short-Term Memory (LSTM) network. LSTM networks actually allow us to solve long term dependency. The mean of learning is weight updating. When weights are updated, they reach a point where the weights are saturated and minor changes happen in new weights. That is why the initial layers don't learn and we call it Vanishing Gradient problem. This problem is related to long sequences because in case of long sequences or long-term dependency this gradient value would be small. If we have a short sequence, then this problem may not be very severe.

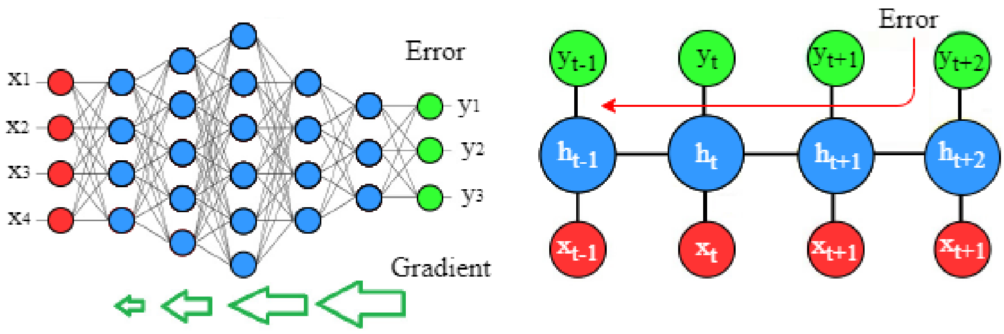

Fig 2.4 Vanishing Gradient a, b [60]

2.5.3 Long Short-Term Memory Networks (LSTM)

LSTMs are also used to evaluate the performance of fake news detection as in [40]. LSTMs are variant of RNN. LSTM is an advanced and sophisticated form of RNN capable of learning long-term dependencies. There are multiple representations of RNN, but the representation shown in Fig.2.5 is intuitive to understand we are revisiting this in LSTM. The blocks in Fig.2.5 are cells of RNNs, where cell with x_t is current time step cell, the block with the input x_{t-1} is previous time step cell and the block with input x_{t+1} is next time step cell. So, x_t is current input if it's a word it can be word embedding, one-hot-vector representation. Here h_{t-1} is a hidden state from 'previous time step cell'. When h_{t-1} comes in 'current time step cell' it gets multiplied by a weight matrix. Current input x_t also multiplied by another weight matrix then after adding the bias $tanh$ is applied. we call it current hidden state h_t.

$$h_t = \tanh\left(W_{hht}\, h_{t-1} + W_{xh}\, x_t + b_h\right) \qquad (2.6)$$

Here x_t and h_{t-1} are concatenated in a single vector. Weight matrix for x_t and h_{t-1} also concatenated into a larger matrix so instead of saying x_t is multiplied by a separate matrix h_{t-1} multiplied by a separate matrix then bias is added and combined these notations wise into a single big matrix which is a weight matrix and a single vector which is having both x_t and h_{t-1}. Even in that representation x_t multiply by its respective weights, h_{t-1} will be multiplied by its respective matrix [61] [62] [63]. General architecture of LSTM is shown and elaborated in detailed in methodology section 3.4.4.

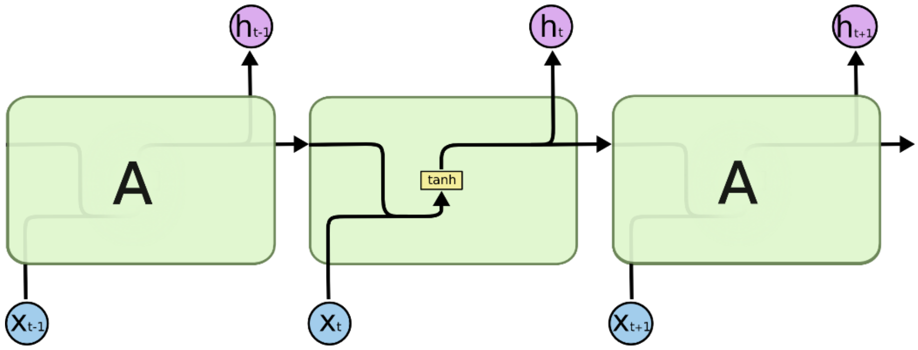

Fig.2.5 RNN revisiting in LSTM [62]

At the end we generally compare ANNs, RNN and LSTMs as the ANNs actually are not able to handle sequential data because they don’t have any memory; they have fixed architecture means the number of inputs and number of outputs is fixed. RNNs are useful for the input data which is sequential or temporal. RNNs can exploit the dependencies in the input and short-term dependencies can learn to predict easily. BRNNs are really just putting two hidden layers or two independent RNNs of opposite directions to the same output. BRNNs are actually used where the context of the input is in different sequences. RNNs basically have local influences and have a problem to solve long term dependency, this is a very common problem in RNNs, and the solution to this problem is Long Short-Term Memory (LSTM) network. These (LSTM) networks actually allow us to solve long term dependency. So, we have also concluded that Long Short-Term Memory (LSTM) is an advanced and sophisticated form of RNN capable of learning long-term dependencies. LSTM can keep track of information passed to it and can update and have a memory unit to save. We input sequential and temporal data to LSTM, its architecture can handle long dependencies in input data but could not solve the issue when we increase the length of the input sentence. Bi-LSTM are just like BRNN putting two LSTMs: one forward pass and the other backward pass.

CHAPTER 3

METHODOLOGY

Fake news detection is a difficult task to be approached by considering all content and social information. This thesis seeks to detect news fakeness based on the news' substance. The main idea behind the detection is to consider the content of news that multiple fact checking websites provided after checking the veracity of the news such as PolitiFact and gossip cop website and then deep learning models to be used for the extraction of latent representation of features that will be used for classification of fake news from true news. The aim of this research is to work on the accuracy performance of the fake news detection by using appropriate dataset based on the content of fake news and deep learning models. Instead of approaching fake news using a knowledge based or style-based approach, we proposed to address fake news content detection using a Latent Based approach. where deep learning models extract latent features automatically without requiring prior knowledge of the field and then classify fake news from real news.

3.1 Proposed Model

Our motivation appears by the fact that current content-based approaches use machine learning framework to represent the content of news in terms of features for classification of Fake news, such as representing the content of news in a knowledge or style-based representation. In most of the research, where fake news is detected from the perspective of content used knowledge or style-based models lacks to produce some promising results in term of accuracy. By reviewing multiple models and techniques in NLP text classification tasks from literature we find out that deep learning models such as CNNs and RNNs advanced model such as LSTMs can be of great use in attaining good features which will help us in yielding promising accuracy. Instead of adopting handcrafting features we developed the idea from Yoon Kim [59] method of extracting features using CNN. Yoon Kin et. al. in [59] developed this idea of extracting features for sentiment classification task by using 1d Convolution. This was the first time when

CNN was used for text related task. All previous works in CNN were related to 2D convolution that were used for the extraction of features from images. These features are then feed to LSTM to classify the fake news from true news. The following diagram shows the workflow of the proposed methodology, see Fig.3.1.

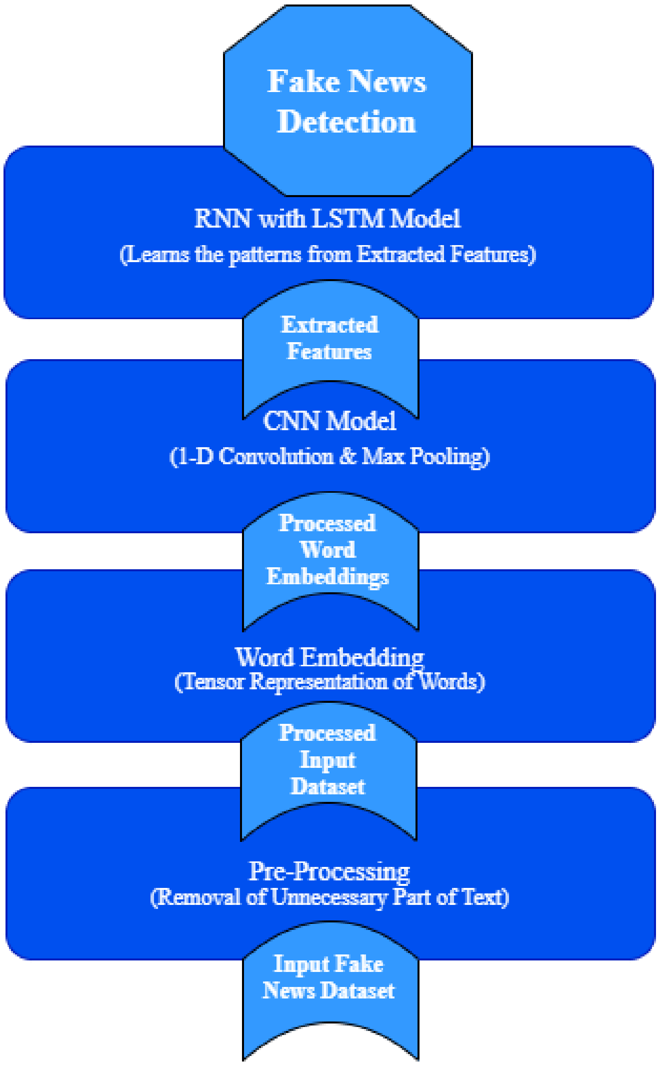

Figure 3.1 Phases of Fake News Detection

Our methodology has distinctive features compared to the traditional Machine learning architectures. Our proposed model does not work on hand crafted features which requires domain knowledge. As in traditional machine learning models features selected by the researchers where one can miss an important feature. In deep learning, Convolution Neural Networks (1d Convolution) can be used as a feature extractor that will extract the most probable accurate features because CNN is designed to learn spatial hierarchies of features automatically and adaptively, from low to high level patterns by using multiple building blocks through backpropagation, such as convolution, pooling, and fully connected layers. Secondly, we have kept the word dependency and memory element in our mind while selecting the methodology, which is an important part in text sequence processing, as the text is a sequence of words, and it requires modelling in a way where coming word should have a connection with the previous and next word. Which gave us the idea to use LSTMs. LSTMs are the advanced form of RNNs which were modeled to overcome the long dependency of the words. The whole process is divided into major phases in developing a Fake News Detection Model as shown in the methodology figure 3.2.

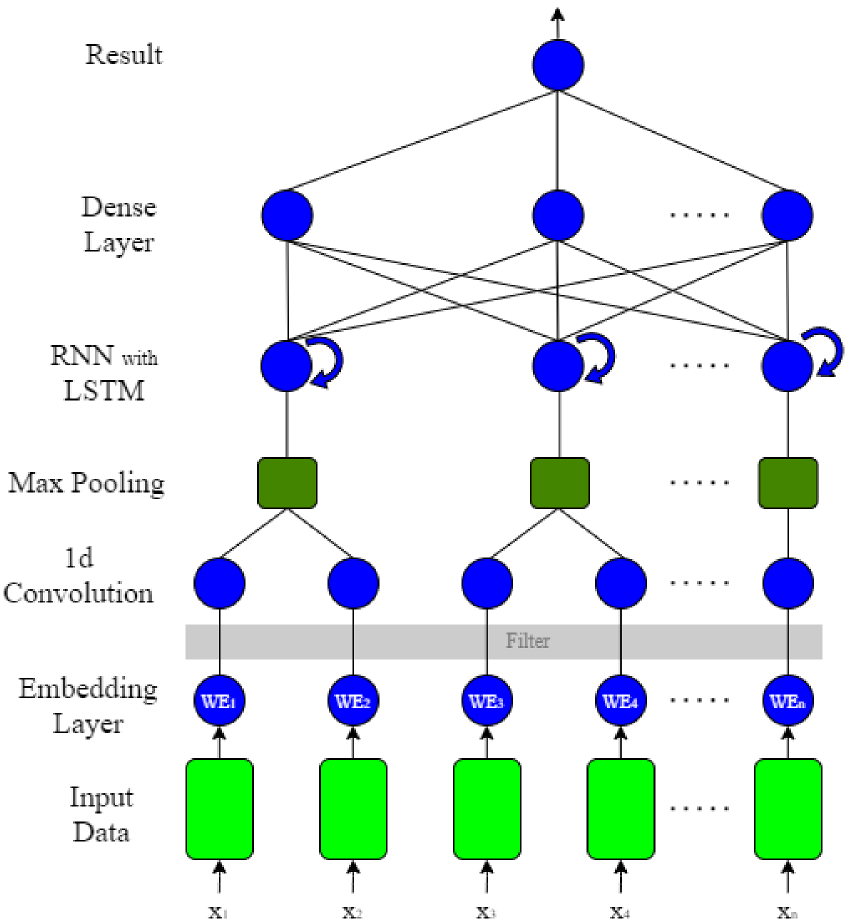

Fig 3.2: Fake News Detection Methodology

3.1.1 Pre-Processing

Preprocessing is an important step towards any technique where the data is transformed to a shape that is required by the author. For example, in our fake news problem, removing stop words, sentence segmentation, tokenization and punctuation removal are required before the data can be represented to embeddings and deep learning models. When the data is cleaned, we will be able to shrink it by deleting unnecessary information. In this context, we removed unwanted features from the datasets. The dataset has some unwanted data. We remove these unwanted or null values because they do not contribute any information to the model. After the inspection of the data, we prepare the data which is called Pre-Processing. In this stage, we convert text to lowercase, and we replace the contractions with their longer forms. The data which is gathered can't be input directly to the Model. To process the data the packages of NLTK and RE are used. Now, we convert contractions into their initial extended forms, for Instance, 'does't' to 'does not'. We have also removed any unwanted characters like "@, #, ?, /". In the descriptions, stop words will only be deleted using NLTK, by deleting these, we are able to train the model faster because there is less data to work with. Then we performed tokenizing of all the words which are in the text. Transforming the tokens into a standard form is the next step once the data has been tokenized.

3.1.2 Word Embedding

A word embedding is a learnt representation for text in which words with the same meaning have a comparable representation. Word embeddings are a type of approach in which individual words are represented as real-valued vectors in a predetermined vector space. Because each word is mapped to a single vector and the vector values are learned in a manner like a neural network, the process is frequently classed into deep learning field. The approach's key characteristic is the use of a densely dispersed representation for each word. A real-valued vector represents each word, which may have tens or hundreds of dimensions. In contrast, sparse word representations, such as a one-hot encoding, require thousands or millions of dimensions. On the basis of words usage, the distributed representation is learned. This permits words which are used in similar manner to have identical representations, capturing their meaning intuitively. This contrasts with the precise but delicate representation in a bag of words approach, where specific words have specific representations regardless of how they are used unless explicitly handled.

Pretrained word embeddings are a more prevalent method. Rather than translating each word to a number, words are translated to tensor representations. Word embedding operates so effectively as it captures the semantics of the words, words with similar meanings have comparable tensor values, and disparities with other word classes are also similar. As can be seen in Figure 3.3, when a word is feminine, its Y value increases. Keras provides an Embedding layer for neural networks using text data. It requires integer encoding of the input data, so each word is reflected by a unique number. Random weights are assigned to Embedding layer is initially and then it will develop an embedding for every word in the training dataset.

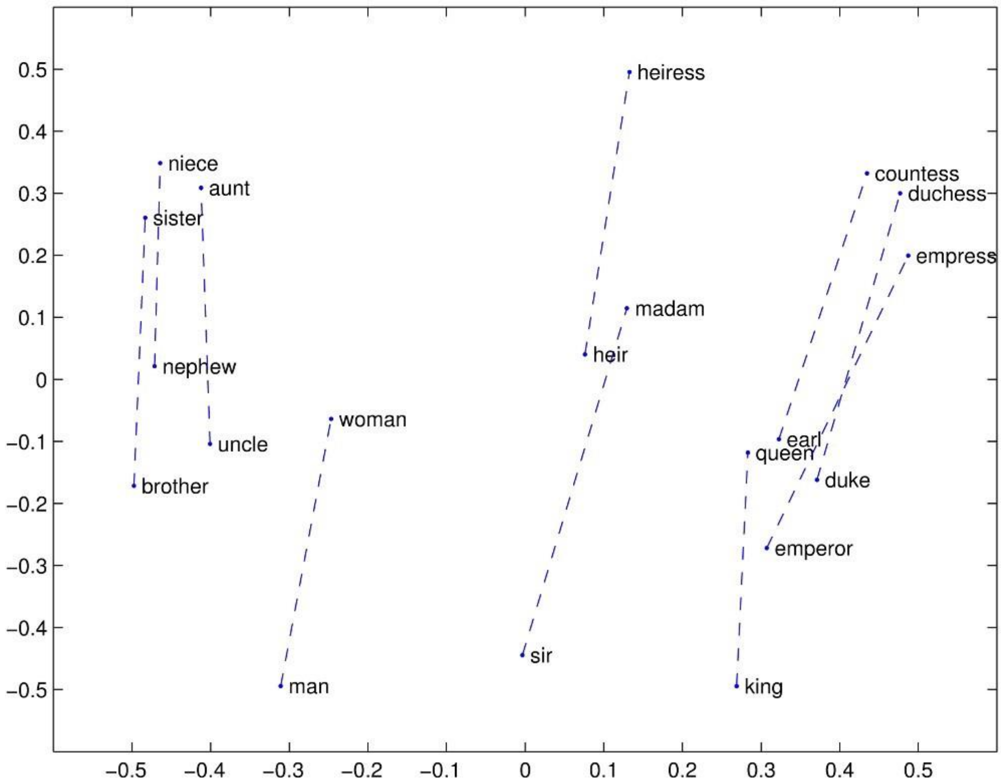

Figure 3.3 Word Embedding - Source: Stanford University

3.1.3 Feature Extraction using 1 Dimensional Convolution

Convolutional Neural Network, In the case of text classification involves a convolutional kernel that works as a sliding window, that captures the embeddings for multiple words, rather than small areas of pixels in an image. Word vectors are used in NLP to represent words in a sentence.

In *1D Convolution*, given a sequence of words $w1{:}n = w1, \dots, wn$, where each is associated with an embedding vector of dimension d. A 1D convolution of width-k is the result of moving a sliding-window of size k over the sentence and applying the same convolution filter or kernel to each window in the sequence, i.e., a dot-product between the concatenation of the embedding vectors in a given window and a weight vector u, which is then often followed by a non-linear activation function g.

Considering a window of words $wi, \dots, wi + k$ the concatenated vector of the i^{th} window is then:

$$x_i = [w_i, w_{i+1}, \dots, w_i + w_{k+i}] \in R^{k \times d} \qquad (3.1)$$

The convolution filter is applied to each window, resulting in scalar values r_i, each for the i^{th} window:

$$r_i = g(x_i \,.\, u) \in R \qquad (3.2)$$

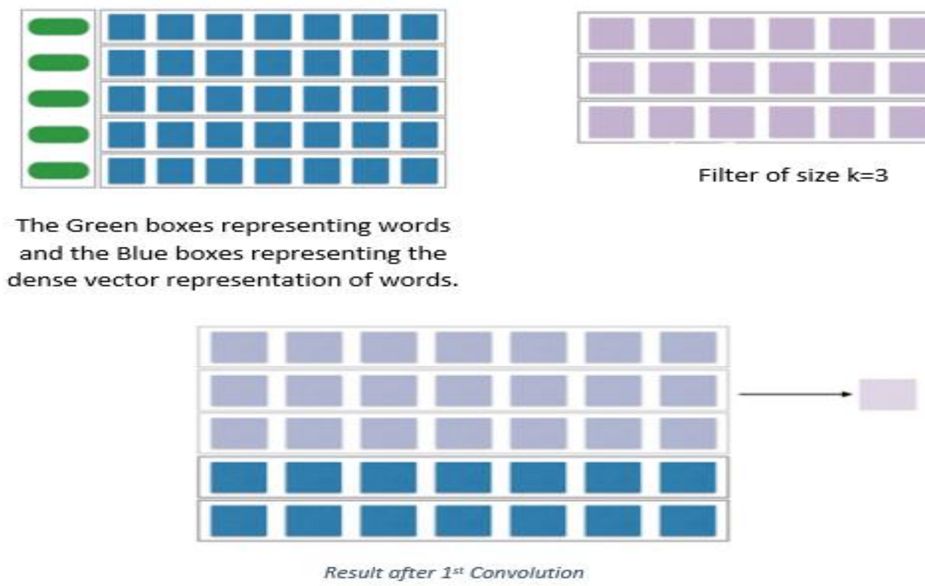

Figure 3.4 Understanding Convolutions in Text, Blink

The example above i.e., Figure 3.4 demonstrates a filter that captures the relationship between two words separated by one word. The filter is multiplied by the word embedding values in each step. The filter values of 1 time the word embedding values result in the word embedding values, while filter values of 0 result in 0.

Once the dot product is taken and a feature map is generated then we extract the most valued features. CNN layer is then followed by the max pooling layer. This layer propagates through the tensors, taking the greatest value each time. The feature space is thus reduced in this manner. This step ensures that vital characteristics or features are retained while empty space is removed, Figure 3.5.

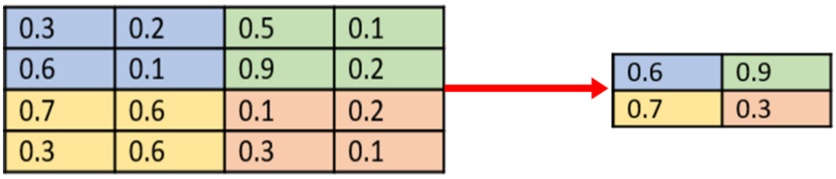

Figure 3.5 Feature Map after Max Pooling Layer

3.1.4 LSTMs Model

LSTMs are expressly intended to avoid the issue of long term word dependency. Retaining knowledge over steps of time is basically their built-in response. All recurrent neural networks take the form of a series of repeating neural network modules. This repeating module in ordinary RNNs will have a single tanh layer which is relatively simple structure. Although LSTMs have a chain like architecture, the recurrent (repeating) module has a different pattern. Instead of a single neural network layer, there are four, each interacting in a unique way. Figure 3.6 transports a full vector from one network node output to the inputs of others. The pink circles indicate pointwise computation such as vector addition, and the yellow boxes represent learnt NN layers. Concatenation is represented by a line merging, whereas forking represents its content being replicated and the copies being sent to other locations. The cell state, which acts similarly to a conveyor belt, is the key of long short term memory.

The LSTM may erase or update information by adding to the cell state, which is precisely regulated via gates. Gates are a way of enabling information to flow through with the option of permitting it to pass through. They're made with a sigmoidal layer and a pointwise multiplication process. The sigmoidal layer produces values ranging from 0 to 1, indicating how much of each element should be allowed through. A value of

0 implies "let nothing through," whereas a value of 1 signifies "allow everything through!"

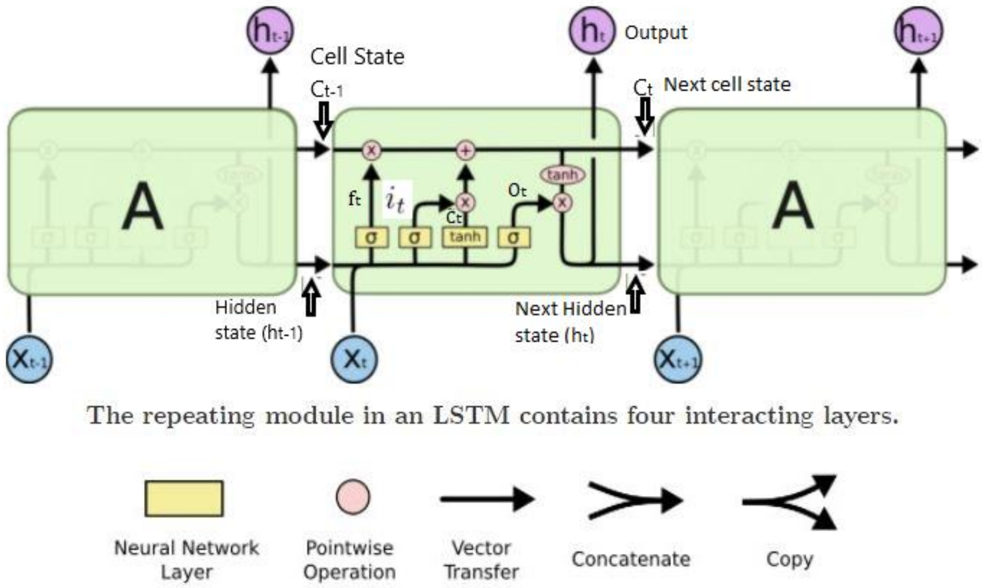

Figure 3.6 LSTM Model [62]

Basically, we have different Gates in the LSTM cell. Gates are implemented using a sigmoid function, sigmoid gives us output between 0 and 1. If an information is multiplied with sigmoid giving 0 it means we are closing the Gate we are not letting any information pass. If sigmoid is close to 1, it means the Gate is open and all the information passes. Since, we have sequences in input $'x_t'$ which means input at time step t. x_t is the input we feed at multiple time steps. Then it is having some information from the previous hidden state which is h_{t-1}, C_{t-1} is the cell state which is the memory of the cell and C_t is a new cell state. The first thing we do is we have a Gate which is called 'forget Gate' we represent it by f_t. It examines h_{t-1} and x_t, and returns a value between 0 and 1 for each integer in the cell state C_{t-1}. A 1 denotes "totally keep this," while a 0 denotes "absolutely get rid of this."

$forget\ layer$

$$f_t = \sigma(W_f.[h_{t-1}, x_t] + b_f) \qquad (3.3)$$

In the second step, we must select what new information we will keep in the cell state. This is divided into two parts. First, a sigmoid layer known as the $input\ gate\ layer$ determines which values will be updated. Following that, a $tanh$ layer generates a vector of new input values, $\tilde{C}_t$, that could be included to the state. We'll merge these two in the next step to generate an updated state of an input Gate. In the third step, we update the old cell state, and, in the end, we get output by output gate.

$$i_t = \sigma(W_i.[h_{t-1}, x_t] + b_i) \tag{3.4}$$

$$\tilde{C}_t = tanh(W_C.[h_{t-1}, x_t] + b_C) \tag{3.5}$$

In the third step, it's time to transition from the old cell state C_{t-1} to the new cell state C_t. The preceding phases have already determined what we should do; all that remains is for us to carry it out. We increase the former state by f_t, forgetting what we agreed to forget previously. After that, we add $i_t * \tilde{C}_t$. This is the updated set of candidate values, scaled by the amount by which we opted to change each state value.

$Cell\ state$

$$C_t = f_t * C_{t-1} + i_t * \tilde{C}_t \tag{3.6}$$

Finally, we use a sigmoid layer to determine which elements of the cell state to output. The cell state is then sent through tanh and multiplied by the sigmoid gate's output, resulting in just the parts we want to output.

$Sigmoid\ gate\ Output$

$$O_t = \sigma(W_o.[h_{t-1}, x_t] + b_o) \tag{3.7}$$

$Output\ h_t$

$$h_t = o_t * tanh\ (C_t) \tag{3.8}$$

The news in dataset are embedded into embedding vectors which are then passed to the our models first layer to use these embedding vectors for further processing. The embedding are then fed to 1 dimentional convolutional neural network to extract the features from the embeddings. Once the features are extracted using 1d convolution and then compressing the features space by Max pooling layer. In which, we have reduced the feature map space by pooling the best feature by considering 2 features. And reduced the feature map. These features are then fed to LSTMs which processess and learn textual long term word dependencies from feature map. We kept the number of LSTMs as blocks

200 to perceive sentenes having length of two hundred words or less as the news on social media or mostly not that long and two hundred words are enough for the news.

we have presented our LSTM module of the proposed model as a many to many problem that learns the long term dependency of the words at each time step, and the information is stored in the cell state and then presented to forget layer as described above it totally pass the information of discard it from the cell state. Our news content is presented to LSTM in form of one word at each time step. And the model learns the embedding of the words and store it in cell state and when new words comes their relation to new words are also learned by the model from the cell state as the cell state is upddated after new information is passed. Once our LSTM model learns the pattern and long term dependencies of the news pieces. The output of the LSTMs after passsing through the activation function is passed to the dense layer to get desired dimentional vector. This layer works as a neural network layer and is fully connected to the previous layer of our model i.e., LSTM layer. In this layer we we get a vector of 64 artificial neurons, which compact the vector space to 1 at the next layer using the sigmoid as an activation function.

Sigmoid function perform the evaluation of the input received if the value is at 0.5 or greater it considers it as the true class giving 1 at the last node of the network and if the value is less than 0.5 and greater than or equal to 0 is outputs it as fake class giving 0 at the output node. As our problem is binary classification problem where we classify the news between being fake or true. So, we have used cross-entropy also known as logrithmic loss as our loss function. Every calculated probability and the actual class output i.e, 1 or 0 is then compared, and their result is computed that castigate the probability relying on variation inbetween the conventional and actual values. The logarithmic penalty, in which less variation showing with a low score probably less that 0.5 and a dominant deviations showing grater that 0.5 or near to 1. Less variations shows the realiabilty of the model, therefore loss of cross-entropy is reduced. The cross entropy or log loss of a framework that describes perfect probability is 0.0. The average cross entropy throughout all samples in our dataset is used to determine cross-entropy for our problem.

CHAPTER 4

EXPERIMENTS AND RESULTS

4.1 Performance Metrics

We employed confusion metrics to evaluate the performance of our proposed approach. The confusion matrix serves as the foundation for most of them. A confusion matrix is a tabular depiction of a classification model's performance on a test set that includes various parameters such as true positive, true negative, false positive, and false negative. (See Table 4.1).

4.1.1 Accuracy

Accuracy is a popular metric that represents the percentage of accurately predicted observations, whether true or incorrect.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \qquad (11)$$

A good model selection is represented by high accuracy; however, because our proposed model is trained for prediction. a news/article in this case predicted as true when it was false positive can undergo negative consequences; likewise, if an news content having factual data is predicted as false, vulnerabilities can arise. As a result, we employed three additional metrics that account for the erroneously classified observation, namely precision, recall, and F1-score.

Table 4.1: Confusion Matrix

	Predicted false	Predicted true
Actual true	False negative (FN)	True positive (TP)
Actual false	True negative (TN)	False positive (FP)

4.1.2 Recall

The total number of positive classifications out of true class is referred to as recall. It refers to the amount of articles anticipated to be true out of the total number of true articles in our example.

$$\text{Recall} = \frac{TP}{TP + FN} \tag{12}$$

4.1.3 Precision

The precision score, on the other hand, reflects the ratio of true positives to all events anticipated to be true. In our scenario, precision is the number of news tagged as true out of all positively anticipated (true) articles:

$$\text{Precision} = \frac{TP}{TP + FP} \tag{13}$$

4.1.4 F1-Score

The F1-score is a measure of the trade-off between precision and recall. It computes the harmonic mean between the two. It considers both false negative and false positive observations. The F1-score can be determined using the formula below:

$$\text{F1} = 2 * \left(\frac{Precision * Recall}{Precision + Recall}\right) \tag{14}$$

4.2 Experiments Procedure

We evaluated our proposed framework on three mentioned datasets, with the intention of predicting if the news articles are true or fake. We divided our experiments into steps and started by preprocessing the data to reduce the data by removing unnecessary information present in the data. For preprocessing, We designed a general processing function for each document to eliminate non-letter characters and punctuation, and then we lowered the letter case in the document. In addition, splitting the sentences to words is used. After splitting by words, Stemming is utilized, which essentially means converting words back to their original form and reducing the amount of word types or groups in the data. We used the Porter stemming algorithm. In the last stage we removed stop words from the data that includes articles, prepositions and conjunctions and some pronouns. After the preprocessing step, the preprocessed data is converted to a tensor representation using $Keras$ embedding layer. we kept the maximum sentence length to 100 and used pre padding.

Once our data is converted to word embeddings, before feeding this data to deep learning models for extraction of features and to get train, we split the data using train-test split into 70% for training and 30% of the data for testing purpose. In order to train our proposed model, we settled some of the hyperparameters for the training purpose, such as embedding vector to 100, sentence length. We kept the number of epochs hyperparameters to 20 epochs. We have used this value so that our model becomes fully trained. Before starting the training process the values of hyper-parameters must be defined, these values express the model's layer size and decides how the model is being trained. Our model has been built with hyper-parameters that are defined with Batch size = 64, Epochs = 20, we have kept the number of filters to 32 and kernel size of 3 in the 1d convolution layer and max pooling window size to 2, we kept the number of neurons in the LSTMs layer to = 100. A complete overview of the parameter is given in table 4.1.

Table 4.2: Complete overview of the model's parameter

```
Model: "sequential"
_________________________________________________________________
Layer (type)                 Output Shape              Param #
=================================================================
embedding (Embedding)        (None, 100, 100)          500000
_________________________________________________________________
conv1d (Conv1D)              (None, 100, 32)           9632
_________________________________________________________________
max_pooling1d (MaxPooling1D) (None, 50, 32)            0
_________________________________________________________________
lstm (LSTM)                  (None, 100)               53200
_________________________________________________________________
dropout (Dropout)            (None, 100)               0
_________________________________________________________________
dense (Dense)                (None, 64)                6464
_________________________________________________________________
dense_1 (Dense)              (None, 1)                 65
=================================================================
Total params: 569,361
Trainable params: 569,361
Non-trainable params: 0
_________________________________________________________________
```

These defined hyper-parameter values are optimal but, in some cases, these values may not be optimal. Hence along with the tuning of these sets of hyper-parameter values, we have obtained optimal results, this procedure is called hyper-parameter tuning

[64]. We find the loss with an optimizer which reduces the loss function, 'Adam' is an optimization algorithm for Stochastic Optimization [65]. We reduce optimization with back propagation techniques. In back propagation we update previous weights. The calculated error is sent back via hidden layers in the time steps. For all the training examples the cost function which we calculate is binary cross entropy also known as Logarithmic loss function. We get the gradients by the calculation of the loss function. The optimizer uses these gradient results to minimize the loss function. Adam optimizer [65] is better than the Stochastic Gradient Descent SGD [66].

4.3 Software/Tool

Training of the model takes too much time; it depends on the hardware. For implementing and training our model, we used Google Collab with GPU settings. To train our model on GPU (Graphic Processing Unit) or TPU (Tensor Processing Unit) is the best idea as we know training over CPU consumes too much time. Google Collab provides the fastest processor GPU or TPU and RAM, it has trained our model. The hardware specification of the lab is 2vCPU @ 2.2GHz, 13GB RAM, 100GB Free Space, Idle cut-off 90 minutes and Maximum of 12 hours. Once the model is trained and generates the output according to the requirements it must be stored because when required this model. We also save our time as the model takes a lot of time during the training. The mode can also be saved in the format of HDF, it is always the best idea to save.

4.4 Dataset

We have used 3 datasets for fake news detection for our experiments. These datasets are: FakeNewsNet: FakeNewsNet is a project that is working on collection of data for fake news research. PolitiFact and BuzzFeed two website of news veracity verification is used from which headlines and news body texts are collected. Social engagements information on twitter is also collected of those articles. All mentioned news content and social context characteristics are included in FakeNewsNet, along with trustworthy ground truth fake news label statistics of the dataset is given in Table 4.3 [27]. An overview of the dataset showing fake news with label 0 and true news with label 1 of BuzzFeed and PolitiFact is shown in fig 4.1 and fig 4.2, respectively.

ISOT Fake news dataset: Our second dataset is completely collected from sources of real world [28] as statics shown in table 4.4. For real news articles, news stories from Reuters.com are collected. The fake news was gathered from a kaggle.com dataset dedicated to fake news. The data set's collector gathered bogus news articles from untrustworthy websites, which PolitiFact (a fact-checking group in the United States) has been working with Facebook to eradicate. There are 44,898 articles in the dataset, with 21,417 representing true and 23,481 as false news. The corpus as a whole includes news from several topics, but political news is the most prominent. Because political news stories are currently the primary target of spammers, we chose to concentrate our efforts solely on them. In this category, both fake and true news pieces were published in the same year, in 2016. Each article has a length of more than 200 characters. Examples from the dataset are shown in figure 4.3, illustrating fake news with label 0 and true news with label 1. (iii) our third dataset is FA-KES, it is based on fake news collected throughout Syrian war, for statics of dataset see table 4.5. To ensure a balanced dataset, this dataset contains news pieces from a variety of media outlets, including mobilization and loyalist press, and print media. The news pieces in this dataset were labelled fact checking using a semi-supervised approach. Human contributors are asked to extract precise and easy-to-extract information that helps match a given article to information representing "ground truth" acquired by the Syrian Violations Documentation Center with the use of crowdsourcing. The outcome is a carefully annotated dataset with 804 articles categorized as true or false, which is perfect for training models to predict news article reliability [29]. Some of the examples from the datasets are given in figure 4.4.

Table 4.3: The statistics of FakeNewsNet Dataset

Foundations	PolitiFact	BuzzFeed
Users	23,865	15,257
Engagements	37,259	25,240
Social Links	574,744	634,750
Candidate news	240	182
True news	120	91
Fake news	120	91
Publisher	91	9

	id	title	text	url	authors	publish_date	label
0	Real_1-Webpage	Another Terrorist Attack in NYC…Why Are we STI...	On Saturday, September 17 at 8:30 pm EST, an e...	http://eaglerising.com/36942/another-terrorist...	View All Posts,Leonora Cravotta	{'$date': 1474528230000}	1
1	Real_10-Webpage	Donald Trump: Drugs a 'Very, Very Big Factor' ...	Less than a day after protests over the police...	http://abcn.ws/2d4lNn9	More Candace,Adam Kelsey,Abc News,More Adam	NaN	1
2	Real_11-Webpage	Obama To UN: 'Giving Up Liberty, Enhances Secu...	Obama To UN: 'Giving Up Liberty, Enhances Secu...	http://rightwingnews.com/barack-obama/obama-un...	Cassy Fiano	{'$date': 1474476044000}	1
3	Real_12-Webpage	Trump vs. Clinton: A Fundamental Clash over Ho...	Getty Images Wealth Of Nations Trump vs. Clint...	http://politi.co/2de2qs0	Jack Shafer,Erick Trickey,Zachary Karabell	{'$date': 1474974420000}	1
4	Real_13-Webpage	President Obama Vetoes 9/11 Victims Bill, Sett...	President Obama today vetoed a bill that would...	http://abcn.ws/2dh2NFs	John Parkinson,More John,Abc News,More Alexander	NaN	1

	id	title	text	url	authors	publish_date	label
0	Fake_1-Webpage	Proof The Mainstream Media Is Manipulating The...	I woke up this morning to find a variation of ...	http://www.addictinginfo.org/2016/09/19/proof-...	Wendy Gittleson	{'$date': 1474243200000}	0
1	Fake_10-Webpage	Charity: Clinton Foundation Distributed "Water...	Former President Bill Clinton and his Clinton ...	http://eaglerising.com/36899/charity-clinton-f...	View All Posts	{'$date': 1474416521000}	0
2	Fake_11-Webpage	A Hillary Clinton Administration May be Entire...	After collapsing just before trying to step in...	http://eaglerising.com/36880/a-hillary-clinton...	View All Posts,Tony Elliott	{'$date': 1474416638000}	0
3	Fake_12-Webpage	Trump's Latest Campaign Promise May Be His Mos...	Donald Trump is, well, deplorable. He's sugges...	http://www.addictinginfo.org/2016/09/19/trumps...	John Prager	{'$date': 1474243200000}	0
4	Fake_13-Webpage	Website is Down For Maintenance	Website is Down For Maintenance	http://www.proudcons.com/clinton-foundation-ca...	NaN	NaN	0

Figure 4.1 True and Fake Examples of BuzzFeed dataset

id	news_url	title	tweet_ids	label
politifact14984	http://www.nfib-sbet.org/	National Federation of Independent Business	967132259869487105\t967164368768196609\t967215...	1
politifact12944	http://www.cq.com/doc/newsmakertranscripts-494...	comments in Fayetteville NC	942953459\t8980098198\t16253717352\t1668513250...	1
politifact333	https://web.archive.org/web/20080204072132/htt...	Romney makes pitch, hoping to close deal : Ele...	NaN	1
politifact4358	https://web.archive.org/web/20110811143753/htt...	Democratic Leaders Say House Democrats Are Uni...	NaN	1
politifact779	https://web.archive.org/web/20070820164107/htt...	Budget of the United States Government, FY 2008	89804710374154240\t91270460595109888\t96039619...	1

id	news_url	title	tweet_ids	label
politifact15014	speedtalk.com/forum/viewtopic.php?t=51650	BREAKING: First NFL Team Declares Bankruptcy O...	937349434668498944\t937379378006282240\t937380...	0
politifact15156	politics2020.info/index.php/2018/03/13/court-o...	Court Orders Obama To Pay $400 Million In Rest...	972666281441878016\t972678396575559680\t972827...	0
politifact14745	www.nscdscamps.org/blog/category/parenting/467...	UPDATE: Second Roy Moore Accuser Works For Mic...	929405740732870656\t929439450400264192\t929439...	0
politifact14355	https://howafrica.com/oscar-pistorius-attempts...	Oscar Pistorius Attempts To Commit Suicide	886941526458347521\t887011300278194176\t887023...	0
politifact15371	http://washingtonsources.org/trump-votes-for-d...	Trump Votes For Death Penalty For Being Gay	915205698212040704\t915242076681506816\t915249...	0

Figure4.2 True and Fake Examples of PolitiFact dataset

Table 4.4: The Statistics of ISOT Dataset

News	Size (Number of articles)	Subjects	
		Type	**Article Size**
Real/True News	21417	World Related News	10145
		News Related To Politics	11272
		News of Government	1570

Fake/False News	23481	News from Middle east	778
		US Related News	783
		left Wingers News	4459
		Politics	6841
		News	9050

	title	text	subject	date	label
0	As U.S. budget fight looms, Republicans flip t...	WASHINGTON (Reuters) - The head of a conservat...	politicsNews	December 31, 2017	1
1	U.S. military to accept transgender recruits o...	WASHINGTON (Reuters) - Transgender people will...	politicsNews	December 29, 2017	1
2	Senior U.S. Republican senator: 'Let Mr. Muell...	WASHINGTON (Reuters) - The special counsel inv...	politicsNews	December 31, 2017	1
3	FBI Russia probe helped by Australian diplomat...	WASHINGTON (Reuters) - Trump campaign adviser ...	politicsNews	December 30, 2017	1
4	Trump wants Postal Service to charge 'much mor...	SEATTLE/WASHINGTON (Reuters) - President Donal...	politicsNews	December 29, 2017	1

	title	text	subject	date	label
0	Donald Trump Sends Out Embarrassing New Year'...	Donald Trump just couldn t wish all Americans ...	News	December 31, 2017	0
1	Drunk Bragging Trump Staffer Started Russian ...	House Intelligence Committee Chairman Devin Nu...	News	December 31, 2017	0
2	Sheriff David Clarke Becomes An Internet Joke...	On Friday, it was revealed that former Milwauk...	News	December 30, 2017	0
3	Trump Is So Obsessed He Even Has Obama's Name...	On Christmas day, Donald Trump announced that ...	News	December 29, 2017	0
4	Pope Francis Just Called Out Donald Trump Dur...	Pope Francis used his annual Christmas Day mes...	News	December 25, 2017	0

Figure 4.3 True and Fake Examples of ISOT dataset

Table 4.5: The Statistics of FA-KES Dataset

Dataset	**Domain (Syrian war)**	**News**	**Number of rows**
FA-KES Dataset (804)	Sources: Mobilization Press, Loyalist Press, Diverse Print Media.	True News	426
		Fake News	378

	unit_id	article_title	article_content	source	date	location	labels
0	1914947530	Syria attack symptoms consistent with nerve ag...	Wed 05 Apr 2017 Syria attack symptoms consiste...	nna	4/5/2017	idlib	0
1	1914947532	Homs governor says U.S. attack caused deaths b...	Fri 07 Apr 2017 at 0914 Homs governor says U.S....	nna	4/7/2017	homs	0
2	1914947533	Death toll from Aleppo bomb attack at least 112	Sun 16 Apr 2017 Death toll from Aleppo bomb at...	nna	4/16/2017	aleppo	0
3	1914947534	Aleppo bomb blast kills six Syrian state TV	Wed 19 Apr 2017 Aleppo bomb blast kills six Sy...	nna	4/19/2017	aleppo	0
4	1914947535	29 Syria Rebels Dead in Fighting for Key Alepp...	Sun 10 Jul 2016 29 Syria Rebels Dead in Fighti...	nna	7/10/2016	aleppo	0
5	1914947536	Suicide bombing kills at least 16 in northeast...	Tue 05 Jul 2016 Suicide bombing kills at least...	nna	7/5/2016	hasakeh	0
6	1914947537	22 dead in heavy U.S. raids on IS Syria strong...	Sun 05 Jul 2015 22 dead in heavy U.S. raids on...	nna	7/5/2015	raqqa	0
7	1914947538	Suicide bomber kills 4 in Assad clans hometown	Sun 22 Feb 2015 Suicide bomber kills 4 in Assa...	nna	2/22/2015	lattakia	0
8	1914947539	Explosion rocks down town Damascus	Sun 01 Feb 2015 Explosion rocks down town Dama...	nna	2/1/2015	damascus	1
9	1914947540	Damascus explosion due to rocket bomb	Sat 24 Aug 2013 Damascus explosion due to rock...	nna	8/24/2013	damascus	0
10	1915433669	Syrian regime steps up aerial assault on Douma	12 February 2015 Casualties mount in the Easte...	alaraby	2/12/2015	damascus	1
11	1915433670	Hizballah leads regime offensive in southern S...	12 February 2015 Free Syrian Army and Nusra Fr...	alaraby	2/12/2015	damascus	1
12	1915433675	Syrian opposition remains divided	23 February 2015 The Sawaiq al-Rahman Brigade ...	alaraby	2/23/2015	idlib	1
13	1915433682	IS video shows murder of 2 Syrian activists	6 July 2015 So called Islamic State group has ...	alaraby	7/6/2015	raqqa	1
14	1915433683	Syrias Nusra Front stages deadly suicide bombi...	7 July 2015 A suicide bomber from al-Qaedas Sy...	alaraby	7/7/2015	aleppo	0
15	1915433684	Regime troops thwart rebel attack in Syrias Al...	8 July 2015 Regime forces fended off an attack...	alaraby	7/8/2015	aleppo	0
16	1915433685	Ahrar al-Sham leader killed in Syria	15 July 2015 Abu Abdelrahman Salqeen leader of...	alaraby	7/15/2015	idlib	1
17	1915433687	Barrel bombs kill 11 in IS town in Syria	16 July 2015 At least 11 civilians including t...	alaraby	7/16/2015	aleppo	1

Figure 4.4 True and Fake Examples from FA-KES dataset

4.5 Results

Figure 4.5 shows the accuracy of model on the FakeNewsNet datasets. The higher accuracy achieved on FakeNewsNet dataset is 92.7%, achieved by our proposed model. Benchmark algorithms RST and LIWC performed poorer than the Shu-TriFN model in [45]. Our Proposed framework achieved best performance measures compared to the benchmark models. RST and LIWC for the news content-based approaches, it is evident that LIWC outperforms RST indicating that LIWC captured linguistic aspects in news contents better. The encouraging LIWC results show that fake news articles have considerably variations from actual news in terms of selecting words that reflect psychometric attributes. we can observe that TriFN consistently leads the other two benchmarks in terms of all evaluation metrics on both datasets. For example, TriFN achieves average relative improvement of 14.50%, 16.01% on BuzzFeed and 19%, 21.4% on PolitiFact, comparing with LIWC in terms of accuracy and F1 score. It supports the importance to model tri-relationship of publisher-news and news-user to better predict fake news. While seeing the proposed framework, the deep learning models outperforms all three baselines i.e., RST, LIWC and TriFN. if we compare the difference percentage of proposed framework with the TriFN algorithm in term of Accuracy and F1 score, on BuzzFeed dataset, our proposed model achieves improvement of 6.3% and 6%, Similarly by on PolitiFact dataset, our proposed framework outperforms the TriFN model by an improvement of 10.8% and 10.2% in term of accuracy and F1 scores. Figure 4.6 and figure 4.7 depicts the evolution of training and testing dataset validation accuracy and validation loss of the model. As from the graphs we can analyze that accuracy of the model is high but not consistent as the news pieces of Fakenewsnet dataset for content based is considered to be no much in numbers and the loss is very less when the model is validated.

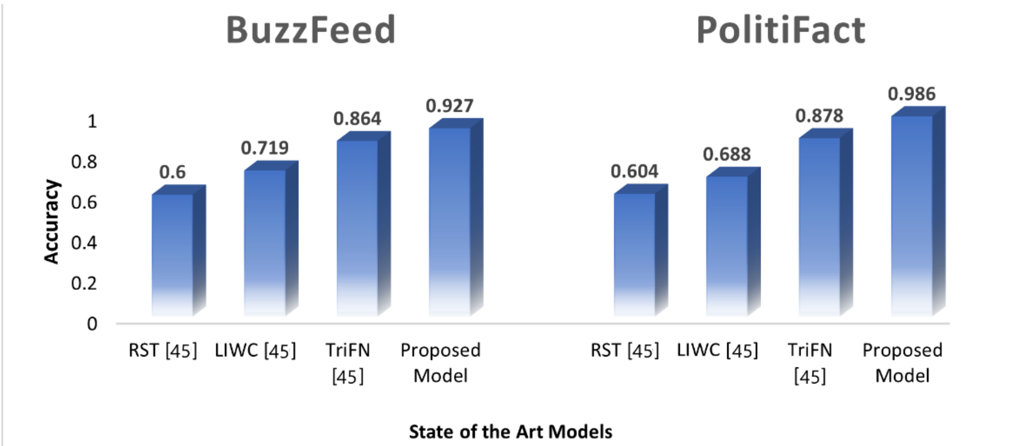

Figure 4.5 Models Performances on FakeNewsNet Dataset

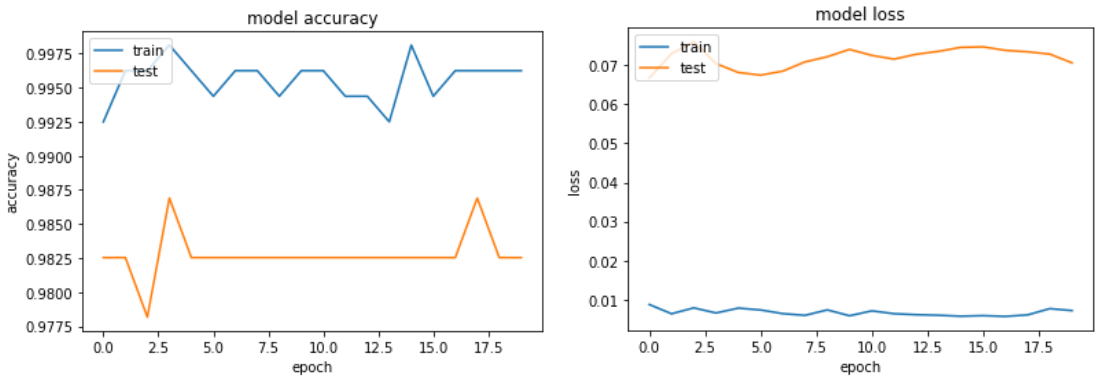

Figure 4.6 PolitiFact dataset test, train validation accuracy and validation loss evolution

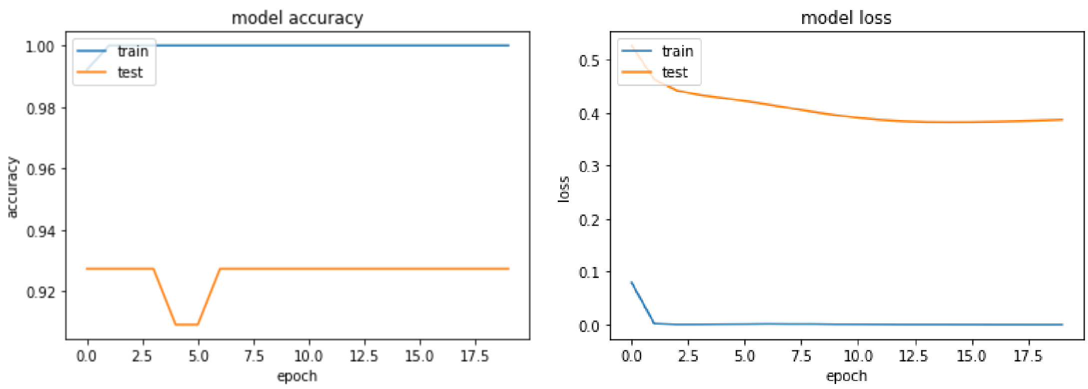

Figure 4.7 BuzzFeed dataset test, train validation accuracy and validation loss evolution

Figure 4.8 depicts proposed framework's performance for classification of fake news in comparison with methods used in [28]. Accuracy of our proposed framework outperforms the models used by the author in [28] by 5.3%, see figure 4.8. which shows that our proposed models achieve better results than the traditional models of Machine learning. Figure 4.9 shows the models training, testing validation accuracy and validation loss. As we can see, the model performed in a very consistent manner the dataset and the model hyperparameter worked well in a compatible mode. The consistency of the results shows that model is best fit on the ISOT dataset. In fig. 4.10, summarize the performance of our proposed framework on FA-KES dataset in comparison with Elhadad et. al. [34]. In [34], author used multiple machine learning models on FA-KES dataset and best accuracy is achieved by Multinomial Naïve bayes model that is 58.09% where our proposed model outperforms the model by an improvement of 31.99%. Figure 4.11 illustrate the evolution of model's validation accuracy and loss. As the high validation loss shows that the model is over fit due to the smaller number of datasets. As the dataset has only 804 entries. This loss can be minimized by increasing the number of examples in the dataset.

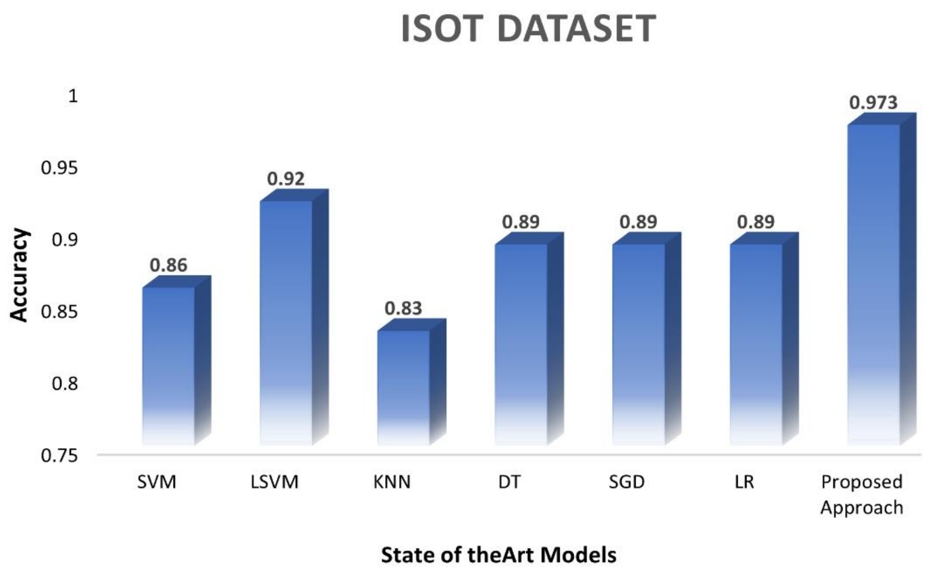

Figure 4.8 Models Performances on ISOT Dataset

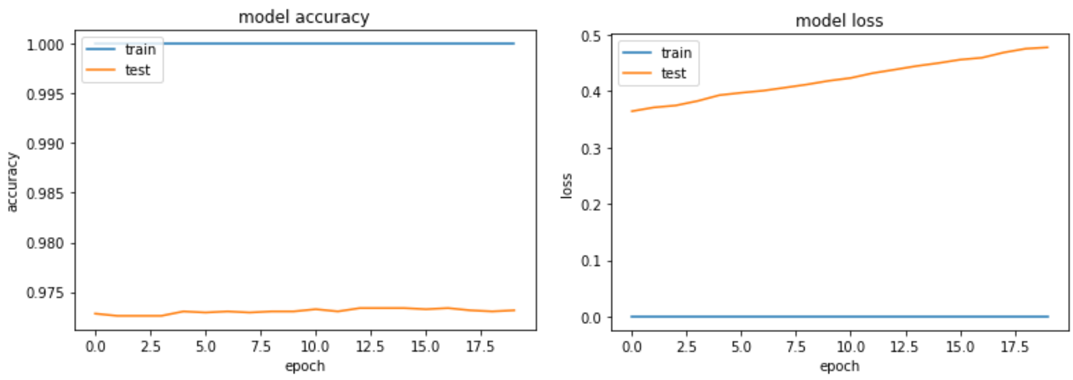

Figure 4.9 ISOT dataset test, train validation accuracy and validation loss evolution

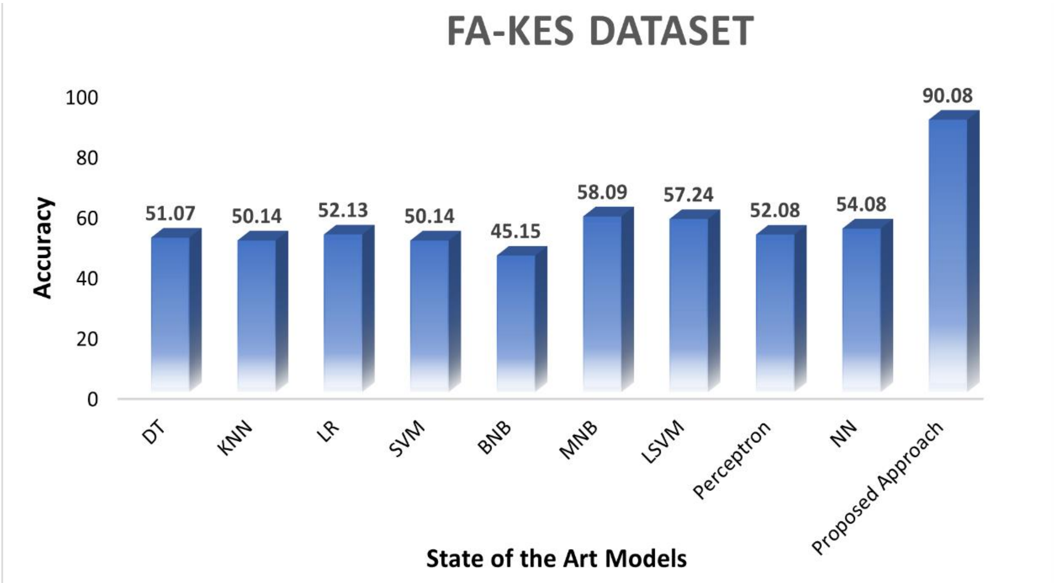

Figure 4.10 Models Performances on FA-KES Dataset

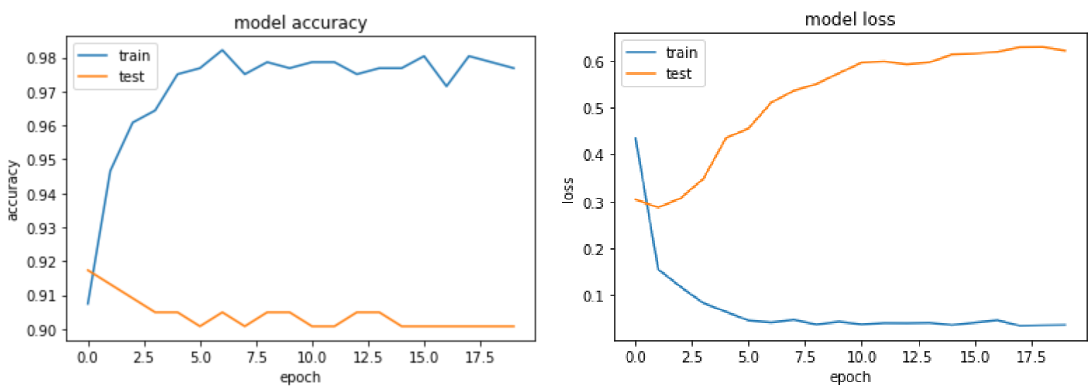

Figure 4.11 FA-KES dataset models accuracy and loss evolution

Results shows that deep learning models used for features extraction and pattern learning with word embedding techniques outperform the handcrafted features and traditional machine learning models. Keeping in mind that word dependencies difficult to be catered in traditional machine learning models which leave quite a gap in achieving well accuracy. As in sequential data it is required to have a connection between the next coming words and previous words in a statement/news. Our proposed approach has LSTM model that has the ability to cater the long word dependencies. This gave our model an edge to learn a better representation of news and classify accordingly.

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Conclusion

In 2016 after the last US presidential elections, the fake news problem gained attention. According to recent research and statistics almost 62% of US adults get news from the social media [12, 13]. Research shows that the most popular fake stories were usually shared on the Facebook rather than the popular mainstream stories [14]. Many people who read the fake news stories have reported that they believe more on fake news than the news from mainstream media. Dewey [15] said that, in US election of 2016, the fake news played a vital role for changing the people's decisions and opinions.

In this research work, firstly we uncover the significance and definitions of fake news. After that we evaluate three benchmark datasets on our proposed model and compare the experimental results with different existing models. Fake news detection is an valuable concept which resolve many serious real-world problems. Despite the fact that the reach of fake news has shrunk, the limitations of automatic disinformation identification remain. The process of fake news detection has important steps that help us to identify and avoiding the spread of misinformation. The social media is an important source of spreading the fake news. To develop the tools for finding and prevention of fake news, it is necessary to understand the human psychology deeply. Machine learning and artificial intelligence are mostly use in existing automatic methods of automatic fake news detection. As well as some image analysis and crowdsourcing methods are also used. With the rapidly increasing of the fake news problem, it is very hot topic for the researcher. The manual fact checking done by professional journalists gives the researchers an opportunity to understand the nature of misinformation and work more efficiently towards the automatic detection of fake news.

In this work we presented a novel approach for the classification of fake news detection based on content of the news. Our proposed framework achieved 92.7% and

98.6% accuracy performance outperforming the state of the art on the FakeNewsNet dataset: BuzzFeed & Politifact and an 97.3% accuracy performance on ISOT dataset and 90.08% on FA-KES dataset by outperforming the state of art models. Our methodology gives a boost in the accomplishment of a better accuracy performance while not needing a lot of training data commonly related with deep learning models.

It is evident from the results of our proposed methodology that by representing the fake news detection problem in latent feature (automatically generated features) deep learning framework where the long term dependencies are also encountered and learned by the model will greatly enhance the model's accuracy and it has showed by our approach by achieving plausible accuracies instead of representing it as knowledge based or style-based feature (Machine learning features).

5.2 Future Work

The complex and dynamic nature of fake news has made it difficult to identifying fake news. After observation, we found several future trends to consider. These future paths are: (i) Fake news early detection requires early recognition of fake news before it spreads and early action to block and intervene. Early detection of fake news is very important because the more fake news spreads, the more people believe it. For early detection of fake news, they mainly rely on information that is limited to news content and social media and face various challenges. First, recent events often provide new and unexpected ideas that are not stored in existing KG or KG or that are difficult to predict. Second, learning to successfully style a bad message in the past may be less useful in the future, especially as deceptive writing styles continue to evolve. Finally, limited information can affect the effectiveness of ML techniques. To address this issue and catch fake news ahead of time, you can focus on keeping the truth up to date. Develop skills related to knowledge-based dynamic (real-time) constructs to summarize basic truths, interoperability and especially misleading writing styles across all subjects, areas of validation, language, and effectiveness. Timely feature updates should be indicated. Valuable content and topics [42] can improve the efficiency of fake news detection. This is explained below. (ii) Trolls and social bot accounts are often catalysts for creating and spreading fake news, which can be very difficult. Hence, further research is needed in the area of social bot detection. The biggest problem is not fake news, but the most dangerous spread of fake news. Sharing fake news with social bots further exacerbates the spread of fake news through the spread of viruses and makes it difficult for experts to automatically detect this content.

References:

[1] Allcott, H. and Gentzkow, M., 2017. Social media and fake news in the 2016 election. Journal of economic perspectives, 31(2), pp.211-36.

[2] Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu. 2019. Fake News: Fundamental Theories, Detection Strategies and Challenges. In The Twelfth ACM International Conference on Web Search and Data Mining. ACM. https://doi.org/10.1145/3289600.3291382.

[3] Victoria L Rubin. 2010. On deception and deception detection: Content analysis of computer-mediated stated beliefs. Proceedings of the Association for Information Science and Technology 47, 1 (2010), 1–10.

[4] Cody Buntain and Jennifer Golbeck. 2017. Automatically identifying fake news in popular twitter threads. In Smart Cloud (SmartCloud), 2017 IEEE International Conference on, pages 208–215. IEEE.

[5] Shu, K.; Sliva, A.; Wang, S.; Tang, J.; and Liu, H. 2017. Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter 19(1):22–36.

[6] M Mitchell Waldrop. 2017. News Feature: The genuine problem of fake news. *Proceedings of the National Academy of Sciences* 114, 48 (2017), 12631–12634.

[7] Yimin Chen, Niall J Conroy, and Victoria L Rubin. Misleading online content: Recognizing clickbait as false news. In *Proceedings of the 2015 ACM on Workshop on Multimodal Deception Detection*, pages 15–19. ACM, 2015.

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

[9] Dan Berkowitz and David Asa Schwartz. 2016. Miley, CNN and The Onion: When fake news becomes realer than real. *Journalism Practice* 10, 1 (2016), 1–17.

[10] Nir Kshetri and Jeffrey Voas. 2017. The Economics of "Fake News". *IT Professional* 6 (2017), 8–12.

[11] Adam Kucharski. 2016. Post-truth: Study epidemiology of fake news. *Nature* 540, 7634 (2016), 525.

[12] Cody Buntain and Jennifer Golbeck. 2017. Automatically Identifying Fake News in Popular Twitter Threads. In *Smart Cloud (SmartCloud), 2017 IEEE International Conference on*. IEEE, 208–215.

[13] Martin Potthast, Johannes Kiesel, Kevin Reinartz, Janek Bevendorff, and Benno Stein. 2017. A Stylometric Inquiry into Hyperpartisan and Fake News. *arXiv preprint arXiv:1702.05638* (2017).

[14] Aditi Gupta, Hemank Lamba, Ponnurangam Kumaraguru, and Anupam Joshi. 2013. Faking sandy: characterizing and identifying fake images on twitter during hurricane sandy. In *Proceedings of the 22nd international conference on World Wide Web*. ACM, 729–736

[15] Xin Dong, Evgeniy Gabrilovich, Geremy Heitz, Wilko Horn, Ni Lao, Kevin Murphy, Thomas Strohmann, Shaohua Sun, and Wei Zhang. 2014. Knowledge vault: A web-scale approach to probabilistic knowledge fusion. In Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM, 601–610.

[16] Carlos Castillo, Marcelo Mendoza, and Barbara Poblete. 2011. Information credibility on twitter. In Proceedings of the 20th international conference on World wide web. ACM, 675–684.

[17] Sejeong Kwon, Meeyoung Cha, Kyomin Jung, Wei Chen, and Yajun Wang. 2013. Prominent features of rumor propagation in online social media. In Data Mining (ICDM), 2013 IEEE 13th International Conference on. IEEE, 1103–1108

[18] Kai Shu, Suhang Wang, and Huan Liu. 2018. Understanding User Profiles on Social Media for Fake News Detection. In 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR). IEEE.

[19] Shuo Yang, Kai Shu, Suhang Wang, Renjie Gu, Fan Wu, and Huan Liu. Unsupervised Fake News Detection on Social Media: A Generative Approach. In AAAI'19.

[20] Robert M Entman. 2007. Framing bias: Media in the distribution of power. Journal of communication 57, 1 (2007), 163–173

[21] Kai Shu, H. Russell Bernard, and Huan Liu. 2018. Studying Fake News via Network Analysis: Detection and Mitigation. CoRR abs/1804.10233 (2018).

[22] Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. *ACM Transactions on Internet Technology (TOIT)*, 17(3):26, 2017.

[23] Zhiwei Jin, Juan Cao, Yu-Gang Jiang, and Yongdong Zhang. News credibility evaluation on microblog with a hierarchical propagation model. In *ICDM'14*.

[24] Zhiwei Jin, Juan Cao, Yongdong Zhang, and Jiebo Luo. News verification by exploiting conflicting social viewpoints in microblogs. In *AAAI'16*.

[25] Xinyi Zhou and Reza Zafarani. 2018. Fake News: A Survey of Research, Detection Methods, and Opportunities. arXiv preprint arXiv:2492706 (2018).

[26] A. Mandelbaum, A. Shalev. Word Embeddings and Their Use in Sentence Classification Tasks. Hebrew University of Jerusalm, (October 27, 2016).

[27] Shu, K., Mahudeswaran, D., Wang, S., Lee, D. and Liu, H., 2018. Fakenewsnet: A data repository with news content, social context and dynamic information for studying fake news on social media. arXiv preprint arXiv:1809.01286.

[28] Ahmed, H., Traore, I., & Saad, S. (2017). 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 (pp. 127{138). Springer.
[29] https://zenodo.org/record/2607278#__sid=js0
[30] Allcott, H., Gentzkow, M. and Yu, C., 2019. Trends in the diffusion of misinformation on social media (No. w25500). National Bureau of Economic Research.
[31] Jindal, N. and Liu, B., 2008, February. Opinion spam and analysis. In Proceedings of the 2008 international conference on web search and data mining (pp. 219-230). ACM.
[32] Yang, M., Lu, Z., Chen, X. and Xu, F., 2017, February. Detecting review spammer groups. In Thirty-First AAAI Conference on Artificial Intelligence.
[33] Ramos. Using TF-IDF to determine word relevance in document queries. In Proceedings of the 1st Instructional Conference on Machine Learning, pp.133–142, 2003.
[34] Elhadad, M. K., Li, K. F., & Gebali, F. (2019). A novel approach for selecting hybrid features from online news textual metadata for fake news detection. In International conference on p2p, parallel, grid, cloud and internet computing (pp. 914–925). Springer.
[35] Victoria L Rubin, Niall J Conroy, and Yimin Chen. 2015b. Towards news verification: Deception detection methods for news discourse. In *Proceedings of the Hawaii International Conference on System Sciences (HICSS48) Symposium on Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium, January*, pages 5–8.
[36] James W Pennebaker, Ryan L Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015. Technical Report.
[37] Shu, K., Wang, S., and Liu, H. (2017b). Exploiting tri-relationship for fake news detection. arXiv preprint arXiv:1712.07709.
[38] William Yang Wang. 2017. " liar, liar pants on fire": A new benchmark dataset for fake news detection. arXiv preprint arXiv:1705.00648 (2017).
[39] Y. Yang, L. Zheng, J. W. Zhang, Q. C. Cui, Z. J. Li, P. S. Yu. TI-CNN: Convolutional Neural Networks for Fake News Detection, [Online], Available: https://arxiv.org/abs/1806.00749, August 1–20, 2018.
[40] Hamid Karimi, Proteek Roy, Sari Saba-Sadiya, and Jiliang Tang. 2018. Multi-source multi-class fake news detection. In Proceedings of the 27th International Conference on Computational Linguistics, pages 1546–1557.
[41] Rubin, V. L., Conroy, N. J., Chen, Y. & Cornwell, S. (2016). Fake News or Truth? Using Satirical Cues to Detect Potentially Misleading News. Proceedings of NAACL-HLT, p. 7–17.
[42] Naeemul Hassan, Chengkai Li, and Mark Tremayne. Detecting check-worthy factual claims in presidential debates. In CIKM'15.

[43] Z. Zhou, H. Guan, M. Bhat, and J. Hsu. Fake News Detection via NLP is Vulnerable to Adversarial Attacks:. In Proceedings of the 11th International Conference on Agents and Artificial Intelligence, pages 794–800, Prague, Czech Republic, 2019. SCITEPRESS - Science and Technology Publications.

[44] Kumar, S. and Shah, N., 2018. False information on web and social media: A survey. arXiv preprint arXiv:1804.08559.

[45] K. Shu, S. H. Wang, H. Liu. Beyond news contents: The role of social context for fake news detection. In Proceedings of the 12th ACM International Conference on Web Search and Data Mining, ACM, New York, USA,
pp. 312–320, 2019. DOI: 10.1145/3289600.3290994.

[46] Song Feng, Ritwik Banerjee, and Yejin Choi. Syntactic stylometry for deception detection. In *ACL'12*.

[47] Z. W. Jin, J. Cao, Y. D. Zhang, J. S. Zhou, Q. Tian. Novel visual and statistical image features for microblogs news verification. IEEE Transactions on Multimedia, vol.19, no.3, pp.598–608, 2017. DOI: 10.1109/TMM.2016. 2617078.

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

[49] L. de Alfaro, M. Di Pierro, E. Tacchini, G. Ballarin, M. Della Vedova, and S. Moret, "Reputation Systems for News on Twitter: A Large-Scale Study," arXiv preprint arXiv:1802.08066, 2018.

[50] Saif M Mohammad, Parinaz Sobhani, and Svetlana Kiritchenko. Stance and sentiment in tweets. ACM Transactions on Internet Technology (TOIT), 17(3):26, 2017.

[51] Ruchansky, N., Seo, S. & Liu, Y. (2017). CSI: A Hybrid Deep Model for Fake News Detection. arXiv:1703.06959.

[52] Rayana, S. and Akoglu, L., 2015, August. Collective opinion spam detection: Bridging review networks and metadata. In Proceedings of the 21th acm sigkdd international conference on knowledge discovery and data mining (pp. 985-994). ACM.

[53] Heydari, A., ali Tavakoli, M., Salim, N. and Heydari, Z., 2015. Detection of review spam: A survey. Expert Systems with Applications, 42(7), pp.3634-3642.

[54] Ye, J., Kumar, S. and Akoglu, L., 2016, March. Temporal opinion spam detection by multivariate indicative signals. In Tenth International AAAI Conference on Web and Social Media.

[55] Ke Wu, Song Yang, and Kenny Q Zhu. 2015. False rumors detection on sina weibo by propagation structures. In *Data Engineering (ICDE), 2015 IEEE 31st International Conference on*. IEEE, 651–662.

[56] Federico Monti, Fabrizio Frasca, Davide Eynard, Damon Mannion, and Michael M Bronstein. 2019. Fake News Detection on Social Media using Geometric Deep Learning. arXiv preprint arXiv:1902.06673 (2019).

[57] Xinyi Zhou, Atishay Jain, Vir V Phoha, and Reza Zafarani. 2019. Fake News Early Detection: A Theory-driven Model. arXiv preprint arXiv:1904.11679 (2019).

[58] Wang H, Raj B, Xing E P. On the origin of deep learning. 2017.

[59] Yoon Kim. 2014. Convolutional neural networks for sentence classification. arXiv preprint arXiv:1408.5882.

[60] github.com/ashishpatel26/Andrew-NG-Notes/blob/master/andrewng-p-5-sequence-models.md

[61] J. Cheng, L. Dong and M. Lapata. Long Short-Term Memory-Networks for Machine Reading.
arXiv:1601.06733v7 [cs.CL] 20 Sep 2016.

[62] Understanding LSTMs – Colah's Blog.

[63] Illustrated Guide to LSTM and GRU: A step by step explanation, Michael Nguyen.

[64] J.Bergstra and Y.Bengio. 2012. Random Search for Hyper-Parameter Optimization. Journal of Machine Learning Research 13 (2012) 281-305.

[65] D.P. Kingma and J. Ba. A Method for Stochastic Optimization. arXiv:1412.6980v9 [cs.LG] 30 Jan 2017.

[66] S.Ruder. June 2017. An overview of gradient descent optimization algorithms. Preprint arXiv:1609.04747v2.