Journal of Theoretical and Applied Information Technology
15th July 2022. Vol.100. No 13
© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

A DETECTION OF UNFAIRNESS ONLINE REVIEWS USING DEEP LEARNING

N.DESHAI[1] , Dr B.BHASKARA RAO[2]

[1]Research Scholar, Gitam, Department of CSE, Vishakhapatnam, India

[2]Professor. Gitam, Department of CSE, Vishakhapatnam, India

E-mail: [1]desaij4@gmail.com, [2]bboddu@gitam.edu

ABSTRACT

In Today's internet world, online activities are growing exponentially and generating a tremendous number of online reviews and ratings, which are a valuable source of information for customers primarily associated with the purchase of marketing, selecting a restaurant, finding products, health, and services, etc. Therefore, online reviews are a crucial part of people's everyday decisions on what to buy, where to buy, where to eat, where to stay, which doctors to see, and what to select based on positive, negative, and neutral. Fake reviews not only mislead innocent clients and influence customers' choice, leading to inaccurate descriptions and sales. However, there is still a significant requirement for a survey that can examine and summarize the various methodologies that are now available. This paper summarizes the existing datasets and the techniques they have acquired to represent the task of fake review detection. In addition, it examines the various feature extraction strategies that are currently available. Finally, we discuss the present gaps in this research area and potential coming directions in this field. We analyze and compare two various features extraction strategies and six various machine classification techniques.

Keywords: *Machine learning, Deep learning, Fake review detection, Feature engineering*

1. INTRODUCTION

The current digital world has become a tremendous source of acquiring users' online reviews and ratings. There is a remarkable growth in online purchasing usage. Therefore, online activities are rapidly increasing, which is more beneficial to organizations and potential customers, who know about items or services before purchasing [19][21]. There has been a considerable growth in consumer reviews due to the internet being available to everyone in recent years. Moreover, potential buyers are influenced by online customer reviews in terms of feedback [33], [34]. To notice product critiques on social media, decide whether to purchase the goods or not. As a result, customer reviews provide a vital service to individuals. Furthermore, positive reviews result in significant financial advantages, whereas negative reviews frequently have a negative economic impact [47], [48]. As a result, as customers become more influential in the marketplace, there is a growing tendency to rely on customer feedback to alter organizations by improving products, services, and commerce [52][54]. For instance, when multiple consumers who bought an exact Acer laptop model filed reviews grumbling about the poor display grade, the factory stood encouraged to create a higher screen resolution type of the laptop. The openness with which customers communicate and their studies have presented websites with consumer critiques. Social media (Twitter, Facebook, and so on) enable anyone to willingly broadcast criticism or comments of any firm at any moment, with no responsibilities and limitations. As a result of the lack of limits, certain businesses utilize social media to broadcast their commodities, trademarks, or marts unfairly or to criticize those of their competitors unfairly. For example, assume rare customers who purchased a distinct digital camera left inadequate evaluations about the image quality. These reviews paint a negative picture of the digital camera to the general audience. As a result, the camera maker may hire somebody or a group to publish fake positive reviews and evaluations concerning the camera.

Likewise, the producer may request that the employed individuals publish nasty remarks about competitors' products to promote the company. Fake reviews are those posted by someone who has not faced the examined things [11]. When a faker collaborates with different fakers to accomplish a clear purpose, the collection of fakers is referred to as a spammer collective [11]. As a result, much research has investigated the subject of fake review

© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

detection and its difficulties. One of the most challenging tasks involved with fake review identification is distinguishing between false and genuine reviews. A reliable and robust system of detecting fake reviews is a crying requirement in today's world to purchase products without being cheated by online sites. We did a literature review to specify existing difficulties and potential paths for future study in this field. It covers classic numerical machine learning and deep learning methods, which intention aid investigators curious about fake review identification in selecting the most appropriate machine learning strategy for their needs.

1.0 Most Impacts of This Survey

Online reviews heavily influence consumers' purchasing decisions. But on the other hand, fake reviews can be used by spammers and scammers to promote products that don't exist or to defame rival products to influence social behavior. In this regard, distinguishing between real and fake reviews has become ever more critical. It is not the first-time research on detecting fake reviews has been undertaken. Others [11], [23], [31], [39], [49], [58], [63], [73], [74] summed up the known strategies on detecting bogus reviews. Although, as indicated in Table 1, such polls have had some drawbacks. For example, researchers did not include all bogus comments, including existing datasets and the latest deep learning methods. The most crucial objective of this study is to give extensive, in-depth text, approach, and accessible datasets that will aid ongoing studies and advances in just this field. Finally, for future research, give real-world data and associated gathering techniques. Table 4 also summarizes the dataset's vital facts, such as the building techniques, the total reviews on every dataset, and related research. We examine the precision and reliability of every technique and the best way for identifying bogus reviews. The following approaches are not used in earlier days: convolutional -LSTM, convolutional -LSTM using character-level convolutions (character-level convolutional -LSTM), HAN (convolutional HAN), BERT (BERT Distil), Rob.

The following is just a breakdown of the publication's structure: The existing extracting features approaches are described in Section 2. Section 3 contains thorough explanations of standard datasets and a Brief Summary. Section 4 discusses known techniques for detecting false reviews and the drawbacks of each strategy, such as everyday ML and NN models. The trials with various methods for reviewing spam identification are shown in Section 5. Section 6 explains the present gaps in this field of study and potential future directions. Finally, the conclusion summarises the significant problems affecting fake review identification and essential findings arising from this research in section 7.

1.1 Fake Review Outline

The growth of internet technology leads to online marketing and associated review sites. As each company nowadays has an occurrence across the online marketing, receiving the exact product and service is more complicated. This sign indicates the reputation of the online reviews across the various platforms. Every individual must depend on online reviews to make purchases and financial decisions. The authenticity of online reviews of a specific product or service, on the other hand, will not always be guaranteed. Some businesses and people use the reviews to promote a particular product or brand while degrading the competitors' products or brands. Deceptive opinions, spammy opinions, and spammy reviews are all terms used to describe fake comments, and their creators have been commonly referred to as spammers. Three forms of spammy opinions could be identified, often referred to as fake reviews [4]. Customers who write bad reviews harm the organizations. Positive reviews to encourage manufactured goods/enterprises to have untruthful thoughts. These evaluations are usually fake or dishonest reviews because they are difficult to distinguish by the readout. After all, authentic and fraudulent reviews are very similar [4]. Only those that comment on the brand of the products are referred to as brand reviews.

Non-reviews that are either irrelevant or merely advertising with no genuine opinion. The last two sorts are disrupting spam opinions, and they pose a small risk and can be easily spotted by anybody examining them [4]. We need to explore the following two Yelp Chi real-world open dataset online reviews [8] to express and understand the life of bogus online reviews. The initial analysis is honest, whereas the next is a hoax. "I like this hotel," says the first reviewer. The work is quite pleasant and will make you feel at ease. Great location, great hotel for a night's stay:) "Wow, what a fantastic location to continue." The staff is lovely then delightful. The advantages are reasonable, such as open bike hire. In addition, the building's history (and renovation) is fascinating. "Thank you for making my stay so special.

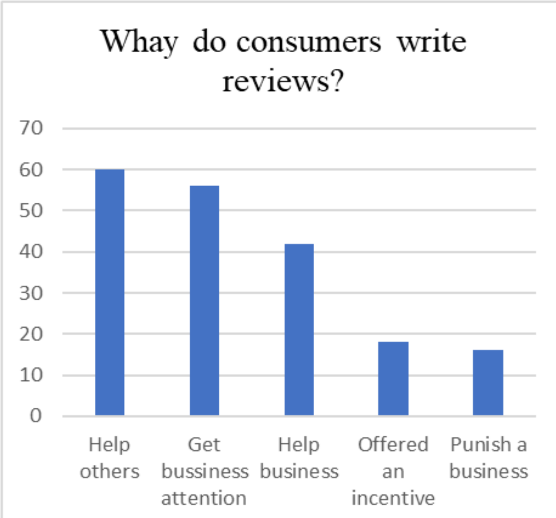

Figure. 1 Online review are help to consumers

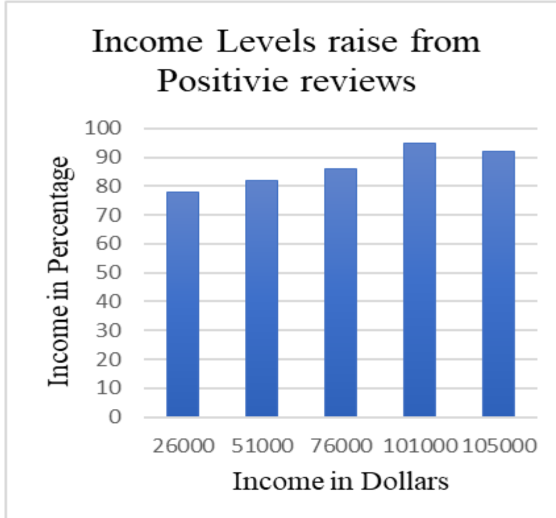

Figure. 2 percentage of Income raise from Online reviews

Figure. 3 percentage of Customer trust Online reviews

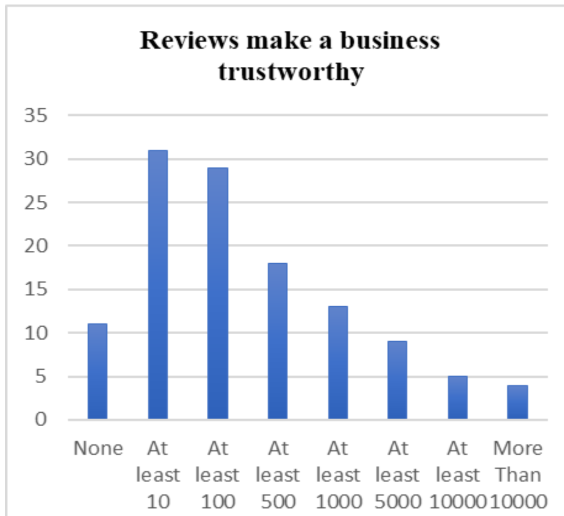

Figure. 4 Online reviews make a business as trust

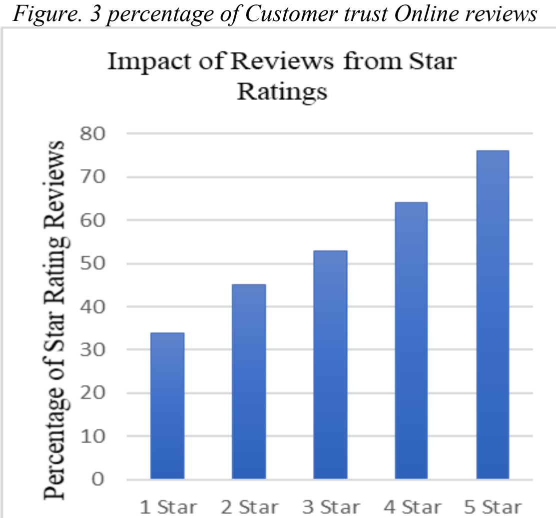

Figure. 5 Impacts Online reviews from Star Ratings

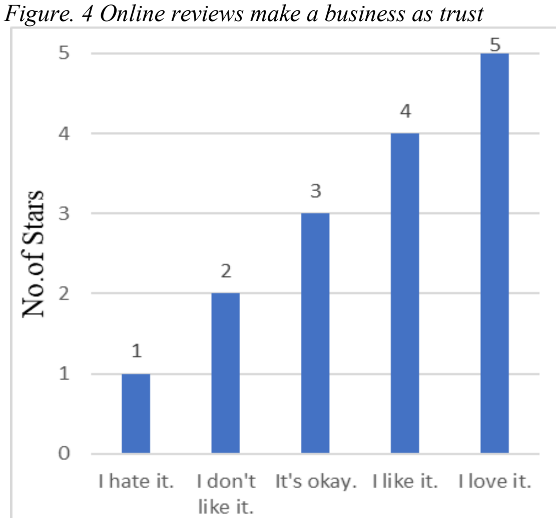

Figure. 6 Best star ratings and meaning of stars

We can conclude that humans find it difficult to discriminate between two reviews simply by reading them. Customers rely heavily on reviews when exploring new firms, and they play an essential role in their decision-making process. Consumers will choose your store based only on the strength of their online reviews. People are more likely to patronize a business if they have received positive ratings, according to 94 percent of respondents as described in figure 3 and figure 4. Moreover because of excellent ratings, people from higher-income categories are even more inclined to employ a company's services, as described in figure 1 and figure 2.

Positive reviews impact all consumers, regardless of their family income. However, we have observed a general pattern of positive reviews having the most significant impact on the purchasing decisions of higher-income groups, is described in figure 5 and figure 6. Just that they are highly related, various scholars physically marked the reviews to categorize them, and their prototype had a 60 percent accuracy rate [77], as shown in Table 1. Consequently, providing effective types to identify fake reviews is usually [73].

Table 1. Identifying Spam Reviews

Year & Author	Dataset & Classifier	Feature Contextual &Measures Precision, Recall, F1
2007 N.Jindal et. al	Amazon, Linear Regression	AUC, Yes
2008 N. Jindal et. al	Amazon, Linear Regression	AUC, Yes
2010 C. Lai et. al	Amazon, SVM	Yes
2010 S. Algur et. al	Web page, SVM	Yes
2011 L.Fangio et. al	Opinions, Linear Regression, SVM Navi Bayes	Yes
2013 M. Arjun et. al	Yelp, SVM	Yes
2014 H. Li et. al	Diapering, SVM	Yes
2014 L.Yuming et. al	Amazon, Linear Regression, SVM	Yes
2016 I. Ahsan et. al	Yelp, NB, SVM	Yes
2016 D. Zhang et. al	Yelp, SVM, DT, RF, NB	Yes

1.2 Tasks for Detecting Fake Reviews

Detection of fake reviews from a great variety of reviews in many areas such as House or Office, Games, and others, each of which has a rating, label CG(Computer Generated Review) or OR (Original Review Generated by People), and review content. The aim is to determine if a review is genuine or fake. It is considered fake if it is computer generated; otherwise, it is not. After product sales, online reviews and ratings could become essential for buying and selling decisions. Fake reviews could affect such choices due to false information, leading to financial losses for the consumers. Identification of fake reviews has thus.

In recent decades, it has received a significant deal of attention. The fundamental challenge with fake review identification is whether spam review detector. A comparison of the most popular feature extraction methods is presented as artificial opinions recognition, identifying the review as false or real in table 1. However, again, machine learning performs a substantial job [73].

For instance, the most common job in fake review identification has supervised learning. Therefore, labeled database false reviews against genuine reviews based on specified characteristics. After reading many of them, it is hard to distinguish between a false and an honest review. Machine learning algorithms could distinguish between bogus and authentic comments by showing underlying language patterns which the human visual misses. A previous study on fake reviews has been classified into three categories detection single spam, a gang of spammers, or fake reviews in only one, mixed, and cross-domain [11]. It is important to note that this study discusses various strategies for detecting fake reviews in NLP. As a result, it focuses primarily on English language evaluations and significant tasks, datasets, and applications

2. FEATURE EXTRACTION

One of the difficulties in text categorization is learning from large amounts of high-dimensional information. Many keywords, words, and phrases appear in documents, resulting in a significant computing strain on the learning process. Moreover, irrelevant or redundant parameters might negatively influence the efficiency and effectiveness of classifiers, as seen in table 2 and table 3. As a result, it is preferable to reduce features to lower the text feature's size and prevent having a vast feature space dimension. Both Term Frequency and TF-Inverted Document Frequency were employed in this survey’s feature selection process. This part examines the existing elements that are being used in the research. These

Journal of Theoretical and Applied Information Technology
15th July 2022. Vol.100. No 13
© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

characteristics could be split into two groups: behavioral and linguistic factors.

Table 2. A comparison of the most popular feature extraction methods

Algorithm, Dataset & Ref.	Dataset Features	Metric & Results
Logistic Regression (LR), Amazon & [4]	Behavioral & review	AUC 78%
LR, Amazon & [4]	Review	AUC 63%
SVM, AMT & [77]	LIC Bigrams	Accuracy 89.8%,89.6%
SVM, AMT & [123]	Deep syntax & Unigram	Accuracy 91.2%
SVM, Yelp & [91]	Behavioral Bigrams	Accuracy 86.1%
SVM, AMT & [99]	N-gram	Accuracy 86%
SVM, AMT & [107]	Stylometric	F1-measure 84%
SVM, AMT & [124]	Unigram	Accuracy 83.21%
Naïve Bayes, Trip-Advisor & [84]	Behavioral & review	F1-measure 63.1%
SAGA, AMT & [100]	LIWC, Unigram & POS	Accuracy 65%
Multi-Iterative Graph-Based Unsupervised Method, AMT & [46]	Behavior, Content, Relation-based features.	Accuracy 95.3%
Logistic Regression & Yelp Chi, NYC, Zip	(Doc2vec) and (Node2vec)	AUC 80.71% 81.29% 83.18%
CNN, AMT dataset [16]	Word2vec word order.	Accuracy
RCNN, AMT & [37]	Word2vec	Accuracy 82.9% 80.8%
Bi-Gated Recurrent Unit with attention, Hotel Restaurant Doctor & [3]	Word Embedding (LIWC, Unigram & POS)	Accuracy 81.3%. 87.01% 76.3%. 83.7%, 57.3%.
Combination of LSTM and CNN, Spam review & [70]	Character-level	Accuracy 99.5%
(Fake GAN), AMT & [7]	Glove2vec	Accuracy 89.2%

2.0 Review-centric features

Size, average word count, sentences per paragraph, percentage of data, and average reviewer sentence length are essential aspects of a review. The percentage of the capitalized word and the percentage of words in each review expresses a positive or negative viewpoint.

2.1 Reviewer-centric features

It provides a detailed review. These features give details on the person. Rather than written material, the review-centric part might focus on reviews. The highest number of reviews that can be done in a day. The proportion of reviews that received excellent or negative ratings. The standard deviation of a reviewer's rating has been based on an average of 10 words.

2.2 Review-Text features

Pretrained neural network models which learn word2vec forms are used to transform every review into a numeric value of 100 elements. 81] [83], Maximum No. of Reviews.:78% of spammers created five reviews per day. 90% of customers write more than one review every day. [84] [94], % Positive Reviews.: Spammer's positive reviews might suggest false reviews with a ratio of only a four or 5-star rating.[72],[81],[88] Average Review Length: Spammers do not offer complete reviews of items, which might help catch them. Also, spammers frequently write reviews quickly since they want to fool 90% of authentic reviewers. [91], Burstiness (Bst): It examines the rate of writing reviews last 24 hours. For fast results, spammers manipulate the ratings. I was writing quickly. [81], [83], [92] Reviewer Deviation: A disinterested reviewer expects the items to be scored by average review rating. Therefore, spammers significantly differ from an item's average ratings. [72] [95], Weighted Rating Deviation: Early variation captures a spammer who spams an assessment quickly after being posted. These spams attract spammers, enabling them to profit from other views. [72], The no of negative reviews about positive reviews: Since a reviewer's positive reviews percentage is crucial, so is their negative reviews percentage. Rate unpleasant reviews with 1 or 2 score ratings. [10] [96] [98], Maximum Content Similarity: The spammer's existence of reviews with similar content. Spammers do the same for other goods. To find out if the author is spamming by comparing the range of their views. [4], Repeated Reviews: No, repetitive reviews for the same item suggest a problem. Due to internet connectivity or operating troubles, the same person may post multiple times.[95], Bottom Ranked Reviews Ratio: Real reviewers rate a product or service before interacting with it, taking more time than spammers, who evaluate quickly to influence consumer decisions. [95], As recently mentioned, posting early reviews suggests that there are false ones. Reviewers' behavior could be questionable when their most recent research is among the best. [83], Extreme Rating Behaviour: Customers could move things from most significant to worst. Likewise, spammers try to away favor or against certain products. [101], First Review Ratio: Initial reviews could hugely affect revenues. Spammers early reviewers to boost their influence and mislead purchasers. Text Features The following are the issue goals: Excellent Feature Extraction Approaches. The weather is posted by an actual client or an artificial intelligence system. Create the most effective DL Classifiers. Meta-data characteristics have demonstrated their utility in detecting fraudulent reviews. Employing meta-data information, unexpected or abnormal reviews could be identified. Whenever a review is recognized as a faker, every review associated with this reviewer could be classified as false. However, such characteristics may not even be present in all data sources, limiting its utility to detect false reviews. For instance, specialists could select specific spam based on the reviewer's identification, such as the reviewer's IP address, review time, and remarks.

2.3 Analyze the following scenarios:

The Internet and online business have grown in popularity in recent years. As a result, there are millions of items and services in online marketing, resulting in an enormous amount of data. As a result, locating the most satisfactory services or goods to meet the need might be difficult. Consumers develop their ideas based on the experiences of others and the feedback they receive from others. Everyone could post anything in today's hyper-competitive market. As a result, the frequency of fake reviews is on the rise. Specific customers occasionally make multiple critical or positive reviews for the same merchandise utilizing a similar computer; this can be considered suspect behavior. When we examine rival product names, we believe the ratings provided by reviewers for each item. We can see that a particular reviewer has written numerous good evaluations for products from that brand.

Additionally, the same reviewer had left countless poor ratings for competitor brands' products. Further, the position of the writer demonstrates the review's value. Finally, it is possible across excellent evaluations made from regions close to the hotel; however, such studies are not legitimate, as the hotel reviewers must be in remote areas.

2.4 Term Frequency (TF)

It is a way of comparing texts that uses word counts from the publications. First, each document could be presented by a vector of the same length containing the text's weights. Then each vector's components are normalized to 1. Finally, I converted each word count into the likelihood of a term appearing in the texts under examination. If the phrase appears in a text, it is written as one; otherwise, it is zero. As a result, each document is a word collection.

2.5 Term Frequency- Inverse Document Frequency (TF-IDF)

It is a weighted measure utilized for extracting information and the NLP mechanism. Generally, it is a mathematical measure used to assess a word's importance in a database. For example, a word's frequency in a document increases its value, yet its frequency in the corpus diminishes its worth. One of IDF's key features is that it devalues standard terms while elevating unusual ones. For example, if we exclusively use TF, phrases like "the" and "then" will defeat the count frequency. Utilizing IDF can minimize the influence of words.

2.6 Part of Speech (POS)

A POS characteristic is the intensity of every POS (Part of Speech) inside the text. Computational linguistics study has revealed that different kinds of studies have different degrees of POS differentiation [102], [103]. Whereas the POS characteristic performs well across domains, it is ineffective to detect fake reviews compared to all other attributes like BoW [77], [100].

2.7 Bag of Words (BoW)

Similarly, such characteristics are recognized as n-gram characteristics and utilized in various NLP responsibilities. Those features display content as a string of phrases or a single phrase.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Several fake review finding techniques have used BoW characteristics such as unigram (1), bigram (2), and trigram (3). On multiple datasets, characteristics related to BoW produce different results [4], [77], [84], [99], [100]. For instance, it achieved 89.6 percent precision when utilizing AMT datasets but only 67.8 percent accuracy when using the Yelp dataset. This is for real-lifetime reviews that differ from reviews gathered through a crowdsourcing platform. The scalability of BoW is a significant limitation (for example, 'this excellent' and 'is this excellent' keep a similar vector interpretation). Furthermore, it is incapable of capturing the semantic meaning of reviews.

Limitation of BoW:

1. Ignores the location information of the word.
2. Ignores the word's semantics.
3. It's essential to have an extensive vocabulary.

2.8 Word Count and Linguistic Inquiry (LIWC)

Text evaluation techniques are widely used and can evaluate semantic characteristics from several perspectives [104] [106]. Aside from unigram, bigram, and trigram features, it's less effective at recognizing fake reviews than most other features [77, 100], which are more efficient at identifying genuine reviews. Integrating n-gram characteristics and LIWC into the classification algorithm, on the other hand, may drastically improve its efficiency. It categorizes into psychological aspects, grouped into four classifications: spoken, personal, linguistics, and psychological characteristics (or psychological elements).

Table 3. Various Features Extraction techniques and their characteristics.

Language Models	Semantics & Syntactical	Context & Out of Vocabulary
Hot encoding	No	No
BoW	No	No
TF	No	No
TF-IDF	No	No
Word2Vec	Yes	No
GloVe	Yes	No
FastText	Yes	No/Yes
Context2Vec	Yes	Yes
CoVe	Yes	Yes/No

2.9 Stylometric.

Syntactic features, as well as word and character-based characteristics, are included [107]. The mean word length and no. of upper-case characters are typically included in word-based, indicating the reviewer's types of surfaces and words. The no. of punctuations used to signify the reviewer's writing fashion is an example of syntactic features.

Features Of Semantic Value. These characteristics of words reveal the purpose or ideas of the terms. These characteristics offer a semantic strategy for identifying bogus reviews [10]. Semantic features outperformed LIWC, POS, and n-grams in cross-domain testing, according to Li et al. [6]. A few years later, Kim et al. [78] introduced an improved classification technique based on Frame Net-based semantic features. They demonstrated that classification performance had greatly improved. On the other hand, these characteristics cannot represent the semantic relationship between records and phrases.

2.10 Word embedding

It is a widely utilized extracting features approach for text data. A low-dimensional vectors representation has been proposed in NLP [108]. Word embedding plus a neural network paradigm generates cutting-edge NLP [109,111]. Compared to word embedding, the standard word vectors technique, where each word could reflect a vector of a FIXED height of the lexicon of the words in the document [112], is far more complex. Word vectors are generated utilizing neural network design, facilitating learning from their environment as described in table 4.

2.11 Word2vec

[113] Predictive methods are commonly used in word2vec. It could be studied utilizing the CBOW method, which forecasts the word centered on its perspective point, or the Continuous Skip-gram technique (CSG), which indicates the closest terms to a particular word [114]. The Skip-gram performs exceptionally well in simplicity and computation efficiency [37].

However, these approaches cannot be learned from terms with few co-occurrences. For unseen words, character2vec (C2V) [115] was proposed to overcome this limitation. Pennington et al. [116] suggested Count-based Glove techniques [117]. This approach does not capture semantic similarity, requires much memory, and does not capture vocabulary words.

2.12 Fast Text

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Fast Text is a library for learning word embeddings and text classification developed by Facebook's Artificial Intelligence Research Lab (AI Research). Embedding could understand character n-gram vectors quicker than word2vec, according to Joulin et al. The researchers [114, 119] presented an unsupervised word vector Pragraph2vec. The skip-gram approach produces sentence vectors for documents, phrases, and paragraphs as described in table 5.

This strategy requires users to learn commonly occurring word groups. This approach is not utilized for data streams since it requires unseen retraining word-groups [120]. Node2vec is an unsupervised learning technique presented by Grover and Leskovec [121].

Table 4. Analysis of word embedding Model

Ref & Data Set	Feature Extraction & Algorithm	Purpose	Result
Paper [189] Movie review,	Unigram & Ensemble Hidden Markov Model	ensemble of text-based hidden Markov models using boostingand clusters of words producedby latent semantic analysis	Accuracy98.1% in Subjectivitydata
Paper [190] Yelp 2013,2014,2015 and IMDB	Skip-gram, CBOW & Text Concept Vector	Text Concept Vector leverages both the NN and the knowledge base to create an advanced quality text presented.	Accuracy71.5% in Yelp 2015
Paper [191] electronic word-of-mouth (eWOM)	TFIDF, Chi & NB	application of text mining techniques to online gender discourse.	Recall 0.65
Paper [192] INEGI and TASS	n-words dan q-grams with or without TF-IDF & SVM	To discover the most effective classifiers, thoroughly investigate all possible combinations of text changes and their parameters.	Best Accuracy 65.4% in INEGI with TF-IDF
Paper [193] 20 Newsgroups, Worldwide Knowledge Base, Reuters	Multivariate Relative Discriminative e Criterion & DT, MNB		Best recall 88.9% Reuters dataset with MNB
Paper [194] World News	Part of Speech, Text Position & KNN	identify whether the contents of news can be exploited for classification or otherwise	F score 0.837 in polarity-based
Paper [195] Suicidality Annotation	Bag of Word n-gram & GA	Suicidality in Dutch-language forum posts is being studied with the help of "supervised text categorization."	F score 0.93
Paper [196] Reuters Corpus Volume 1 (RCV1-v2) contains 804,414 newswires	TFIDF, LIT, LIB term weighting, IG selection & Relaxation Hierarchy	Improve performance classification	F score 0.8436 in TFIDF
Paper [197] Saudi Press Agency contains a set of 1526 Arabic documents	TFIDF, LIT, LIB term weighting, IG selection & Hybrid Associative Classification	increase the efficiency of the classification process	F score 0.856
Paper [198] ACM Digital Library, Reuters 21578	CSI, EB, MF, TFISF, TR & NB, SVM, LR, RF, AdaBoost, Bagging, Dagging, Rs, Majority Voting	examines the predictive performance	F score 0.91 Bagging and Random Forest.
Paper [199] Reuters-21578, 20Newsgroup -18821 documents, Worldwide Knowledge Base	TF, TF-IDF & SVM, K-NK, NB, Extreme Learning Machine	improve the performance of text categorization	F1 micro average 97% from SVM
Paper [200] 20 Newgroups 18846 docs, Reuters R8 7674 docs,	TF-IDF & SVM, NB	Semi-supervised self-training of LDA text categorization using topic model representations	Better in NB
Paper [201] WAP, K1a, K 1b,r,e0, and re1, 20 Newsgroups,	Normalized Difference Measure (NDM) & Multinomial NB, SVM	propose a new feature ranking (FR) called NDM	NDM is 66% better than other
Paper [202] News article in the Indonesian language	TFIDF, SVD & Multinomial NB, Multivariate NB, Gaussian NB, SVM	find the appropriate algorithm to classify a news article in Indonesian Lanthe guage automatically	Recall 0.984 in TFIDF and MNB
Paper [203] Dataset of 16,323 accident records from US OSHA	n-gram & SVM, LR, RF, K-NN, NB	Classification methods for text mining should be evaluated.	Best perform SVM with unigram dan RBF

Journal of Theoretical and Applied Information Technology
15th July 2022. Vol.100. No 13
© 2022 Little Lion Scientific

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Table 5. Text classification methods are compared.

Author(s)	Architecture	Innovation	Feature Extraction	Details	Corpus & Measures	Drawback
B.J. Sowmya et al. [201]	Hierarchical architecture is of Rocchio's work.	One of the hierarchical data classifiers was an innovation.	TF-IDF & Rocchio Algorithm	Evaluate & verify the ranges using CUDA, mostly on GPU.	Wikipedia serves as the corpus & F1-Macro	Only operates hierarchical data sets and finds a few documents.
Bloehdorn et al. [203]	AdaBoost	AdaBoost technique for novelties with linguistic capabilities	Model Boosting technique for Feature Extraction is BOW.	One of the many algorithms used in ensemble learning.	Reuters-21578 & F1-Macro and Micro	Complicated computations and a reduction in readability
A. Genkin et al. [206]	Bayesian LR	LR analysis of data with many dimensions	TF-IDF & Logistic Regression	It's established on the Gaussian distribution. An LR Analysis	The RCV1-v2 corpus and the F1-Macro measure	To make predictions, several different variables must be considered.
S.B et al. [207]	Architecture is a method of increasing the weight of a structure.	TheMultivariate Poisson Model for Text Categorization is a new approach to text classification.	Extraction of Characteristics Weights Words and Model Nave Bayes Kim	To extract the Poisson parameter, we used a per-document TF normalization.	Reuters-21578 & F1-Macro	This creates a big assumption about the data distribution, which is not supported by the data.
K. Chen et al. [208]	Inverse Gravity Moment	TFIGM was introduced.	TF-IDF, TFIGM & SVM and KNN	Integrates a statistical model to accurately quantify a term's ability to discriminate across classes of words.	20 Newsgroups and Reuters-21578 & F1-Macro	It fails to obtain employment and is still semantic, with an unsolved semantic problem.
H. Lodhi et al. [209]	Kernel for String Subsequence's	The application of a certain kernel	Similarity using TF-IDF & Support Vector Machines	A kernel is the inner product of the feature space formed by all subsequences of length k, the feature space generated by the kernel.	Reuters-21578 & F1-Macro	The outcomes were not made transparent, which was disappointing.
T.Chen et al. [210]	BiLSTM-CRF	In classifying data, use an NN sequence model.	Word Embedding & CRF	The categorization can improve in sentence-level sentiment classification of sentence types.	Reviewers' comments and accuracy	It doesn't work with words that have not been encountered before and with high computing complexity.
Z. Yang et al. [211]	Attention Networks with a Hierarchical Structure	It is organized hierarchically.	Word Embedding & DL Techniques	Attention processes are applied at the word and sentence levels accordingly.	Yelp, IMDB, & Amazon Review & Accuracy	It performs for documents only at the document level.
J. Chen et al. [212]	DNN	2-D TF-IDF with CNN	2D TF-IDF & DL Techniques	A new approach to detecting verbal hostility developed.	Twitter Twittes& F1-Macro and Micro	A model architecture that is data-dependent has been devised.
M. Jiang et al. [213]	Deep Belief Network	DBN & SoftMax regression is used to create a hybrid text classification model for text classification.	DBN & Deep Learning	DBN accomplishes the feature learning process to address the HD & sparse matrix issue & SoftMax regression to categorize the text information.	Reuters-21578 and 20-Newsgroup & Error-rate	In terms of computation, this approach is costly, and approach representation remains a challenge with this paradigm.
X. Zhang et al. [214]	CNN	ConvNets for text classification	Encoded Characters & DL	Character-level There are six convolutional layers and 3fully	Yelp, Amazon reviews, Yahoo!Answers data Set & Relative Errors	This model is solely intended to find aspects of their inputs that are position-invariant.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

				connected layers in the ConvNet network.		
K. Kowsari [215]	CNN, DNN and RNN	It provides a solution to determining the optimal DL structure and architecture.	TF-IDF, GloVe & DL	RDML	IMDB review, Reuters-21578, 20NewsGroup, and WOS	Computationally, this is time-consuming.
K. Kowsari [216]	Hierarchical structure	Stacks deep learning DL architectures to deliver knowledge at every document level	TF-IDF and GloVe & DL	HDLTex	Web of science data set & Accuracy	It applies only to hierarchical data sets.

Table 6. The literature contains a wealth of information about freely available datasets.

Domain Datasets & authors	Data Construction Method	Total Reviews (users)	Publications (data used by)	Review Text	Rating	Image	Product	Reviewer
Yelp CHI [8] Restaurants & Hotel	Filtering algorithm	67,365 (38,063)	[1],[22], [14],[36], [42],[50], [57],[65]	Yes	Yes	No	Yes	Yes
Yelp NYC [72] Restaurants	Filtering algorithm	359,052 (160,25)	[44],[44], [43]	Yes	Yes	No	Yes	Yes
Yelp ZIP [72] Restaurants	Filtering algorithm	608,598 (260,277)	[41],[44], [43],[75], [76],[78]	Yes	Yes	No	Yes	Yes
Yelp [79] Consumer Electronics	Filtering algorithm	18,912	[79]	Yes	Yes	No	Yes	Yes
Dianping[80] Restaurants	Filtering algorithm	9,765 (9,067)	[81]	Yes	No	No	Yes	Yes
Amazon [4] DVD, Music, Book	Ruled-based Technique	5,8 million (2,15 M)	[27],[5]	Yes	No	No	Yes	Yes
Amazon [89] Books	Ruled-based Technique	6,819 (4,811)	[29]	Yes	No	No	Yes	Yes
TripAdvisor [90] Hotels	Ruled-based Technique	2,848	[28],[64]	Yes	No	No	No	No
TripAdvisor [93]	Human	3,000	[93]	Yes	No	No	Yes	No
Opinions [84]	Human	6,000	[29]	Yes	No	No	Yes	No
TripAdvisor [77]	Amazon Mechanical Turk	800	[37],[16], [28]	Yes	No	No	Yes	No
TripAdvisor [99]	Amazon Mechanical Turk	1,600	[37],[16], [7],[78]	Yes	No	No	Yes	No
TripAdvisor [100]	Amazon Mechanical Turk	3,032	[4],[27], [32],[37], [12],[3], [29],[65]	Yes	No	No	Yes	No

Word embedding is a frequently utilized approach in NLP. Techniques like Fast Text and Glove improved accuracy and speed in NLP [122].

2.13Glove

A variation on the word2vec approach for effectively learning word vectors, developed by Pennington and colleagues at Stanford, the GloVe algorithm is an extension of this method. The gloVe is an unsupervised learning approach for creating vector representations of words. The University of Michigan developed it.. The gloVe is a technique that combines the global statistics of matrix factorization techniques such as LSA (Latent Semantic Analysis) with the local context-based learning of word2vec to create a hybrid system. A word-context or word co-occurrence matrix is created explicitly rather than implicitly by GloVe, utilizing statistics over the entire text corpus rather than by employing a window to establish local context. Feature extraction is a data extraction approach. The most popular Feature in the fake review detecting area was analyzed. Combining features to train the classifier exceeds utilizing a single feature type [85], [86], [97]. For example, BOW alone is most remarkable to LIWC and POS, according to Mukherjee et al. [97]. Another analysis [91] indicated that behavioral traits outperform linguistic criteria. Thus, integrating behavioral and text characteristics increases the effectiveness of false review detection algorithms.

3. BENCHMARK DATASETS

Table 6 summarizes the literature's existing datasets. Then, as seen in Fig 2, we classify these datasets into four types.

3.0 Method Of Filtering Algorithm

The impact of product reviews on the business platform is growing, giving consumers more information about their products and directly influencing consumers' buying decisions. Mukherjee et al. [8] collected 67,365 Reviews on the yelp of restaurants and motels in Chicago from 2004 to 2012. The Yelp spam filter assessed reviews as genuine or fake. They employed linguistic and behavioral variables to learn classifications. Ads on the website internal data like geolocation, IP address, session logs, and networks captured customer behaviors. Rayana and Akoglu [72] used the same strategy to acquire two more Yelp.com datasets, NYC, and ZIP, from 2004 to 2015. 359.052 reviews on NYC and 608.598 on zipping. In the same way, the Yelp spam filter classified every review. Thus, the average Yelp dataset is 130.6. Later, Li et al. [80] created datasets in Chinese with a review length mean is 85.5. This dataset has 9.765 total reviews. But these datasets were constructed using an undisclosed filtering technique to designate each review as fake or real.

3.1 Method of Human Beings

To test their hypothesis, Li et al. Three undergraduates were requested to write fake reviews. Every student categorized an investigation to assess its authenticity. The majority voting rule, where independent human judges are prejudiced, was used to forecast "fake review." Finally, they obtained the dataset, which included 6,000 reviews, 1,398 of which were deemed to be fake. Ren et al. [93] created a dataset with 3,000 reviews, 712 deemed fake. Manual annotation, on the other hand, necessitates a large workforce. Furthermore, artificial recognition accuracy remains low [99]. As a result, many mislabeled reviews stay on these datasets.

3.2 Method Of the Amazon Mechanical Turk (AMT)

This section's datasets were compiled using crowdsourcing platforms. Massive data collection is possible using crowdsourcing services. It primarily describes the task of the network website and pays for the job to be completed by anonymous online workers. Humanity cannot distinguish between genuine and fraudulent reviews, but they would add them. Ott et al. [77] gathered 800 AMT restaurant evaluations in Chicago. They got 400 natural and 400 fake TripAdvisor reviews. As a result, Ott et al. [99] created datasets with 1,600 reviews, 800 of which were affected. Li et al. [100] used a similar technique to create datasets, which received 3,032 reviews. On the other hand, this data's distribution differs from a real-world dataset.

3.3 Method Based on Ruler

Jindal and Liu [4] used Amazon's rule-based technique to build a dataset. Reviewer IDs on the same product, duplicates on other goods, and various reviewer IDs on other products. Three types of repeated reviews were compared using the Jaccard distance approach. They reject studies with a similarity greater than 0.9. With 5.8 million reviews, 55,000 were deemed fraudulent. Correspondingly, the researchers [89, 90] created datasets comprising 6,819 and 2.848 reviews for the book and hotel domains. Barbados et al. [79] used Yelp.com's web-scraper to crawl review databases. They identified 9653 fake reviews and 20828 authentic reviews based on content and consumer characteristics. These datasets were labeled using the rule-based approach. The rule-based method eliminates the need for costly handwritten annotations. The noise is present in various types of annotation data. They regarded it as fraudulent. Many consumers reviewed the same product, the same impact, or different items. When consumers get many reviews for the same effect because of poor product administration or an internet network, they flag the reviews as false. As a result, we need to discuss this annotation technique.

3.4 Fake Review Detection in Literature Reviews

Nowadays, customers increasingly rely on online reviews for decision-making, and online retailers regard reviews as a norm. Face, speech, font, fraud detection, and disease diagnosis have been solved via ML [125] [131]. [132] [135] in the past several centuries, ML has been utilized to detect fraud in online apps such as SMS, email, and blogs [132]. Here are some methods for spotting fake reviews and some of the advantages and disadvantages of each.

Table 7. Summary of supervised traditional statistical machine learning models in One domain, mixed domain, and cross-domain for fake reviews detection.

Ref & Dataset	Method & Features	Results	Comments
[4] Hotel, Restaurant, and Doctor	Ensemble machine learning Unigram and bigram	Accuracy: NB 87.1% and Ensemble techniques 87.68%	Using the term "fake" to describe several reviews is misleading.
[12] Hotel, Restaurant, and Doctor	Sparse Additive Generative Model (SAGE) LIWC, POS Unigram	Hotel Unigram accuracy:76.1% Restaurant unigram:77% POS:74.6% LIWC:74.2% Doctor unigram:52% POS:63.4% LIWC:64.7%	The suggested model was unable to retrieve the sentence's linguistic features. Fake reviews in a single domain can be detected effectively with the proposed techniques.
[29] Dever dataset spam dataset Opinion`s dataset	Support vector network (SVN) LIWC, WSM, LDA	LDA accuracy:90.9% OpSpam dataset, 94.9% DeRev dataset, 87.5% Abortion dataset, 87% Best Friend dataset80% Death penalty dataset.80% Mix Domain, LDA & WSM accuracy:76.3%. Cross-domain, WSM, and LDA accuracy:59.3% on DeRev datasets and 64% on Best Friend dataset.	A deep neural network is arguably the best way to increase cross-domain effectiveness.
[27] Damping Real-life	Hybrid supervised ML Behavioral and Content.	The accuracy on combination features:98%, Behavioural features:74,% Content features:69%	The findings of this study showed that the behaviors of the reviewer are temporal dynamic.
[15] The gold standard dataset Hotel, Restaurant, and Doctor	Decision tree Feature selection	F-measure: 76.91% on the Yelp dataset Hotel domain F-measure:78.3. Restaurant domain F-measure:81.8% Doctor domain F-measure:75.0%	When picking characteristics, it is vital to consider data correlation. For example, models based on neural networks are superior, whereas the one suggested here is inferior.
[42] Yelp Chi dataset	NB, RF. JRip and AdaBoost J48 classifiers. TFIDF Feature selection.	AdaBoost accuracy:73.4	In several settings, the results were shown to be unstable. It is not enough to compare the suggested model to conventional ML techniques to establish its effectiveness.
[1] Yelp Chi dataset Trip Advisor collection	Naïve Bayes CNN Multinominal Naive Bayes MDL Text RF Rocchio SVM For text representation by TFIDF with N-gram	F-measure on TripAdvisor negative reviews using SVM:87.3% F-measure on TripAdvisor positive:89.9% F-measure on TripAdvisor (negative and positive review):89.9% F-measure on Yelp using MDL Text:71.7%	Over time, the model's efficiency deteriorated. Sentiment polarity has an impact on results. On the Yelp datasets, MDL Text performed much better. The TripAdvisor dataset proved to be the best for SVM.
[18] Yelp Chi & Semi-real dataset.	Ensemble Learning Model TFIDF Feature selection	Ensemble learning F1 measure:81.7%. Ensemble learning F1 measure:76.1% on an artificial dataset	The chi-Squared feature has a significant impact on performance. It is not enough to compare the suggested model to conventional ML techniques to establish its efficiency.
[28] AMT dataset	Ensemble learning model Unigram & bigram features	accuracy NB:87.12% RF :84.87% SVM :87.68% Stacking :87.68% Voting:87.43%	Notably, deep learning algorithms performed better. A review embedding approach employing deep learning can improve results.
[38] Hotel, Doctor Restaurant	Adaption model Character n-gram	Restaurant:79.3%. Doctor:63.8%	Cross-domain detection of fraudulent reviews is challenging; thus, deep neural networks may be more suitable.
[51] Yelp Chi. Yelp NYC Yelp ZIP Yelp Consumer Electronic	Analysis of concept drift (SVM, LR, and PNN) TF-IDF	Accuracy on Yelp Chi:68.17%, Yelp ZIP:91.35% , Yelp NYC:84.85% , Yelp Consumer Electronic:76.72%	The efficiency declined dramatically over time due to the review's shifting features. A high relationship between concept drift and classification performance harms the prediction system.
[60] Dataset collected from Yelp.com	SVM NB RF MLP Lexicon-based method (SentiWordNet)	Acuuracy RF :92.9% SVM:84.9% NB :73.5% MLP :83.6%	Rating sentiment inconsistency characteristics help discover fake reviews. The suggested model uses little data with Word embedding.
[66] Yelp Chi	Ensemble model (RF, Xgboost, Lightbm, Catboost, and GBDT) Review centric Reviewers centric	F1-score using stacking Hotel:72.06%. Majority voting:71.51%. Restaurant:79.46%. majority voting:78.97%	It was stacking outperformed majority voting. However, the modern process did not exceed the proposed model. Also, it is time-consuming.
[1] Yelp Chi dataset Trip Advisor collection	Naïve Bayes CNN Multinominal Naive Bayes MDL Text RF Rocchio SVM Text representation by N-gram with TFIDF	F-measure on TripAdvisor negative reviews using SVM:87.3% F-measure on TripAdvisor positive:89.9% F-measure on TripAdvisor (negative and positive review):89.9% F-measure on Yelp using MDL Text:71.7%	Model performance deteriorated. Polarity influenced performance. Product and service variety affects performance. On the Yelp datasets, MDL Text performed best. On the TripAdvisor dataset, SVM performed best.

4. SPOT FAKE ONLINE REVIEWS WITH CONVENTIONAL MACHINE LEARNING METHODS

Identifying false reviews relies heavily on machine learning, which could be separated into supervised, semi-supervised, and unsupervised learning.

4.0 Statistically Supervised Traditional Learning in Fake Review Detection

Supervised learning algorithms are used to spot bogus reviews. This section summarizes the research on supervised learning (see Table 5). For example, Jindal and Liu [4] used duplicate checks to identify bogus reviews. The suggested model has two elements: The first phase utilized unigram and bigram characteristics and Nave Bayes, Random Forest, and SVM as classification techniques. The second part used two ensemble approaches to increase classification efficiency (stacking and voting). Additional percent Primitive characteristics seem to have a significant impact on efficiency as described in table 7.

Unfortunately, the recommended framework cannot perform effectively on the unbalanced dataset. Inspired by this, Khurshid et al. [18] built on their earlier effort and created an ensemble approach for detecting false reviews depending on specific attributes. Tier 1 employed classifiers (Discriminative Multi-nominal N0042, a library for SVM, and J48), whereas Tier 2 employed an LR classifier to provide the right outcome. Researchers successfully extracted structural and language characteristics using the following feature options: Particle swarm optimization was utilized to examine the attribute set, Cuckoo Search was being used to investigate the attribute space, performed Greedy iterative in vector space, and Chi-Squared was used to calculate the value of the Chi-Squared statistic value to estimate the importance of a characteristic. They tested the suggested model [8] and a semi-real database [77]. The chi-squared attribute considerably increases the offered performance, with a reliability of 84. According to the experimental tests, percent on the restaurant's dataset on yelp and 81.7 percent on the semi-real datasets. Though, incorporating the chi-squared characteristic into the learning algorithm could enhance the suggested performance of the model. Cardoso et al. [1] conducted a comparative overview of multiple content-based classifier techniques to see if information properties shift over time. The models' efficiency declined dramatically with time, according to the experimental data using real-world datasets from Yelp [8]. This is due to spammers' constant attempts to evade the spam filter. Further, the latest studies incorporate traits not even revealed by a classification model in previous reviews in the current world. They also revealed that the models' efficiency deteriorated greatly with time. As a result, novel models are needed to deal with changes occurring in fake review characteristics. Furthermore, the intensity of the reviews impacted the approaches' effectiveness. As a result, they advised employing a different approach for every polarity type. They also discovered that the effectiveness of the methods might be influenced by the variety of items and services available. They suggested having a different paradigm for every service or product. Sánchez-Junquera et al. [139] suggested a character n-gram feature-based fake review detection algorithm. As different classifiers, they used an SVM and NB. The suggested technique was tested on domains including the 'Death sentence,' 'Abortion,' and 'Best Friend' dataset [140]. According to the results, the suggested model outperformed on SVM with LIWC, LDA & words, and Deep syntax & words [138] in detecting bogus reviews. However, when compared to other methodologies [29], [123], [141], the outcomes became worse. This shows that adding a characteristic mixture to a classifier can help it function better. Mani et al. [28] proposed a supervised learning approach for detecting false reviews depending on unigram employing the ensemble method, and the bigram characteristics approach had two parts. Random Forest, Nave Bayes, and Support Vector Machine have utilization techniques during the first study. In the second stage, assaulting and polling ensemble approaches are developed to increase the categorization. According to the experimental tests, the NB had the highest accuracy, 87.21 percent in the first step, mostly on the gold standard dataset [99]. On the other hand, the stacked ensemble approach outperformed votes with only an accurate 87.68 percent framework, demonstrating the relevance of employing an ensemble approach to identify false reviews. On the other hand, the suggested framework did not outperform DL techniques. Alizadeh et al. [142] suggested a false review detection technique based on textual and metadata elements resulting from prior work. [81] proposed a hybrid supervised machine learning strategy for detecting spammers. They discovered reviewers' behavior; hence they presented a labeled Markov approach for identifying spam based on a

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

single reviewer posting time. Then, using Co-bursting, they modified the approach to include multi-hidden Markov to determine posting signals, and the behaviors are bursting technique to aid in the detecting spammers. They employed a real dataset from Dianping [80], although these techniques did not apply any measures to estimate the classifier model. Sánchez-Junquera et al. [38] suggested an adaptation model identified across the study. The recommended model domain characteristics using Co-occurring Entropy and then mask them using a wrong approach. The suggested model had trouble recognizing cross-domain based on the gold standard dataset finding utilizing naïve utilizing classifiers. The authors [51] pointed out a problem with concept drift in fake online reviews whenever the qualities of reviews alter over time. The authors used authorizing learning techniques, and benchmark concept drift detection methods to investigate and illustrate their point. The authors used statistical ML techniques, and benchmarking theory drift finding approaches to examine and illustrate their point. They looked at four real-world datasets in Yelp [8, 72] and [79] and discovered that the classifier's performance declined dramatically as false review attributes changed over time. They also discovered a strong link between idea drift and classifier performance, hurting the forecast algorithm's performance. This research emphasizes the review methods to address this problem. Also, the authors [60] created a methodology problem. Also, the other authors [60] created a methodology for investigating inconsistencies in fake reviews identification depending on various criteria (language, text, and ratings). The retrieved features are input into various ML classifiers (SVM, NB, RF, and MLP) to determine whether an online review is fraudulent or authentic as described in table 8 and table 9. To test the proposed model, they gathered data from Yelp.com. The experiments prove that review discrepancy features could improve the detection of fake reviews. The suggested model only works with limited data, and combining deep learning with word embedding representation can improve performance. Yao et al. [66] suggested an ensemble model for detecting fake reviews based on review content and characteristics using a combination approach and resampling to deal with the unbalanced data, determining the appropriate sampler portion for every classifier. The collected characteristics are then supplied to every classifier independently. Lastly, they used popular voting and stacking procedures to improve the classifier model's execution. The suggested technique couldn't perform approaches in experimental findings on the Yelp dataset [8]. In addition, the suggested technique has time complications.

4.1 Conventional Statistical Unsupervised

Learning In Detecting Fake Reviews Supervised learning is not always suitable based on the complexity of creating adequately labeled datasets. This section highlights the unsupervised learning strategies currently available in the works, as indicated in Table 6. Since unsupervised learning does not require labeled data, it could tackle this problem. [10] Lau et al. Researchers have proposed an unsupervised technique and p Semantic Language Model to detect fictitious customer reviews. The suggested approach was based on Jindal ad Liu's [4] premise that two duplicate reviews also Fake checks were identified using the similarity measure algorithm, which was then personally confirmed. The thoughts that did not have a cosine resemblance with every other review beyond a particular threshold, on either hand, were preserved as accurate reviews and were not physically evaluated. The Amazon.com dataset includes 54,618 reviews, with 6% flagged as false. The SLM technique has been used to assign a spammy rating to every thought. The suggested model's experimental findings achieved an AUC score of 0.9987, outperforming SVM. SLM was also successful at identifying reviews that were not genuine. However, assuming that all the reviews on a product are fraudulent is not always accurate. Dong et al. created an unsupervised sentiment technique [30] to detect fake online reviews afterward. The suggested model technique has four layers: word, document, subject, and sentiment. To find the reviews' opinions on a topic, they improved the LDA technique, which was utilized to gather data from documents. RF and SVM classifiers are provided with sentiment and subject characteristics. As well as sentiment and content, the Gibbs sampling procedure [144] was employed to derive the probability distribution over the various subjects and phrases. The proposed document-level model outperformed existing models with features such as POS, LDA with unigrams, and N-grams on the real-world dataset from Yelp.com. This model was only evaluated with content-based methods, which was inadequate for determining its efficiency, neglecting the reviewer's behavioral characteristics. In response to this, Li and his colleagues came up with the idea of using a list of selected themes to identify bogus reviews [59]. They then employ K-means can do the reviews based on their corresponding groupings and subjects. Finally, they employed material duplication and time bursts to classify a suspect group as fake. The fake performed well in experiments using data gathered from JD.com. Consideration of duplicate content as fraudulent, on the other hand, is suspect. Therefore, the unsupervised graph-based model provided by Noekhah et al. [46] was used to detect fake reviews utilizing its characteristics to extract the text's semantic meaning. Based on the crowdsourced and AMT datasets, it was found that using a combination of features improved the effectiveness of the fake review recognition paradigm [77]. A comparison with other models of neural networks did not reveal that the proposed model was any more effective.

Table 8. Models for identifying fraudulent reviews in one, mixed, and cross-domain contexts are summarized.

Ref	Dataset & Features	Method	Results	Significant Outlines
[10]	Amazon reviews & considering duplicate studies as fake	Unsupervised model and eloped SLM.	AUC:87%	Outperformed SVM SLM effectively detected fake reviews Cons. ering duplicate reviews as fake are unreliable.
[15]	Yelp.com Hotel, Restaurant, and Doctor & Feature selection method	Decision Tree method.	F-measure: Yelp: 76.91% Hotel: 78.3% Restaurant:81.8% Doctor:75.0%	Consideration of data correlation in selecting appropriate features can lead to greater efficiency.
[30]	Yelp & Latent Dirichlet allocation (LDA)	Unsupervised topic sentiment standard common probabilistic method.	F1 measure Restaurant:83.92% Hotel:85.03%	LDA's performance can be improved by including behavioral elements. The effectiveness of the proposed approach could be enhanced by integrating it with research [40].
[46]	ATM dataset from Amazon & Content, Behaviour, Relation-based features.	The method is based on an unsupervised, iterative graph.	Accuracy ATM data:95.3% Crowdsourced:93%	When compared to a single model, the combining features provided good efficiency. An iterative technique and network structure could improve the commission.
[59]	JD.com & LDA	Unsupervised learning	Accuracy:96.42%	duplicate content considers fake & unreliable.

Table 9: Fake reviews can be detected using semi-supervised statistical machine learning algorithms.

Ref & Dataset	Method & Features	Results	Comments
[9] Gold standard dataset	Novel PU method (MPIP-UL) Latent Dirichlet Allocation	Accuracy:79.2%	The proposed model outer performed previous PU learning models
[13] The gold standard dataset: Hotel, Restaurant, Doctor.	Multi-task method (MTL-LLR) Unigram and bigram features	Accuracy on Doctor:85.4% Hotel:88.7% Restaurant:87.5%	An issue with cross-domain review detection checking is possible by applying transfer learning. [26]
[35] JD.com	PU semi-supervised learning Review content features Metadata features	Accuracy: On 200 test data with 600 training data is 87.6% and on 100 test data with 700 training data is 89.3%	The suggestion standard did not execute well on the short text (less than 20 words)
[43] Yelp Chi Yelp NYC Yelp Zip	Semi-Supervised learning framework (SPR2EP) Doc2vec, Node2vec	AUC on Yelp: Chi:80.71% NYC:81.29% Zip:83.18%	The suggested model performs with long text only.
[55] Yelp Chi AMT	Ramp one-class SVM TF-IDF	Accuracy: AMT 92.3% Yelp Chi:74.34%	Outliers and noise could influence the OC-SVM classifier's decision function in the proposed framework.
[61] Yelp CHI	Semi-supervised (SVM, NBRF, LR, KNN, LDA, and DT) Review text Reviewer features	Co-training multi fusion features Precision:83.97% Recall:84.45% F1-Score:81.89%	Including reviewer, abilities could help. They utilize deep learning to construct a robust and efficient fake review detection approach.
[68] AMT dataset Yelp dataset	Investigated the effectiveness of semi-supervised learning method Bigram	Accuracy: Co-training on the AMT:88% Self-training on AMT:93% TSVM on the AMT:83% Yelp:69% Self-training on the Yelp:73% TSVM on the Yelp:64%	The metadata content about the reviews could increase effectiveness.

© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

4.1 Conventional Statistical Semi-Supervised Learning In Detecting Fake Reviews

Because of the difficulties in getting labeled reviews, semi-supervised learning recognizes as false reviews; the research regarding semi-supervised learning approaches has been summarized in Table 7. It's used frequently in text categorization with good performance [145, 146]. However, to identify false reviews, Yafeng et al. [9] suggested unique n-labeled learning and Positive

Approach termed blending population and particular nature PU learning.

We gained valuable positive and negative instances with latent Dirichlet, and K means integration. The Dirichlet procedure mixing approach grouped such fake reviews into various categories. Finally, researchers combined individual and population techniques to detect bogus review groups. The final classifiers have been constructed utilizing MKL. The suggested approach beat earlier PU learning methods in terms of reliability. For example, Deng et al. [35] proposed a PU semi-supervised learning algorithm to recognize fakes reviews. First, researchers used the review's similarities (duplicate or close duplicate) to identify fakes. Next, researchers used the K-Means algorithms to sort the reviews into groups based on the number of bogus reviews. They assigned a specified threshold to every category. They classified a review positively since it's not recognized unfortunate case. Appraisals are generally bad if they closely match the genuine negative case. They obtained the data from JD.com.

The suggested technique identified fake reviews with an overall accuracy of 88.1 percent. But the suggested framework failed to execute effectively with sentences under 20 words. Researchers cannot compare their approach to other alternatives to establish its efficiency. Hai et al. [13] developed a multi-task approach (SMTL-LLR) to identify bogus reviews. Laplacian Regularized Logistic Regression (LRLR) has been utilized using the unlabelled data approach semi-supervised multi-task technique (SMTL-LLR). The suggested framework of motivation and learning for a specific task applies the skills from some other similar operation. They chose 10,000 unlabeled reviews from Ott et al. [77]'s dataset (Doctor, Hotel, and Restaurant). In several domains (Doctor, Hotel, and Restaurant), SMTL-LLR beats state-of-the-art approaches [4], [77], [147] [149]. This information can help enhance the categorization performance of the model. For example, the latest research by Yilmaz and Durahim [43] used textual information and reviewer items networking properties to identify bogus reviews. We employed [119] and [121]'s unsupervised learning techniques (Doc2vec and node2vec). Node2vec produced node embedding through network data, whereas Doc2vec produced text embedding from review content. A reviewer object functionality has been created by linking things (hotel and restaurant). We are using node2vec to establish the generative model between objects and reviewers. Then a logistic regression model classifies the review as spam r not. Researchers tested the model across 3 Yelp datasets [8], [72]. Just on three datasets, the suggested framework using merged characteristics beat the state of art approaches [72], [150] by 80.7percentnt, 81.2percentnt, and 83.18percentt, correspondingly. Node2vec outperformed Doc2ve.

Furthermore, the proposed method was not matched to all other techniques like a neural network. Due to the lack of labeled datasets, a semi-supervised process termed 'Ramp One-Class SVM' was subsequently used to detect false reviews [55]. The suggested model performed well on the AMT dataset, achieving 92.3% efficiency on the Yelp dataset, gaining 74.37 percent accuracy. But the suggested framework could not beat existing approaches. The researcher [61] proposed detecting fake reviews depending on review content and reviewer attributes in some other research. They suggested measuring how much emotions could boost an achievement as a first step. Their second strategy involved using unlabeled data and continuously combining the training data to modify derived characteristics. Their next step was to employ seven machine-learning algorithms to determine if the review was fraudulent or not. The suggested algorithm performed well on the Yelp dataset regarding accuracy and recall. Ligthart et al. [68] recently evaluated their efficacy. Semi-supervised review fraud detection. They employed self-training, co-training, and Trans inductive SVM (TSVM) [151]. The AMT and Yelp Chi datasets indicated that self-training with a classification Algorithm performed much better on both datasets.

Predictive values are learned using traditional machine learning methods. It is also easy to accomplish and requires low computation power. Moreover, different conventional machine learning methods have been driven by deep learning models with various datasets. Therefore, feature extraction is a complex process that requires knowledge from the entire dataset. It also performs poorly on massive data compared to deep learning.

© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

5.0 DETECTING FAKE REVIEWS WITH NEURAL NETWORK (NN)

Natural language processing tasks can benefit significantly from neural network models [114, [152] [155]. Deep learning techniques, such as most fair representation neural network models, could rapidly obtain valuable attributes compared to conventional machine learning. The linguistic interpretation of a bit of text could likewise be captured utilizing only a word embedding approach in deep understanding. Recurrent neural networks, convolutional neural networks, and long short-term memories have been used to identify fake reviews (LSTM).

5.1 Detecting Fake Reviews with Convolutional Neural Network (CNN)

In computer vision, CNNs are a particular type of neural network. The local features critical for categorizing NLP tasks could be captured using CNN. In Fig. 3, we demonstrate a small analogy of a CNN technique to detect review spam as showed in table 11. The input review's word embeddings will be first divided into a matrix. This matrix is supplied in the convolution layers, which comprise various filter of multiple dimensions. Transfer of outcomes from convolution
layers to max-pooling is step two. Finally, combine the pooling outcomes to determine the final representative vector. The last vector predicts the review label. Table 8 summarises the various CNN approaches proposed in the literature and will be discussed in this section. Convolutional Neural Networks were used by Li et al. [6] to initiate a neural network model for learning document representation designed to check misleading spam opinions. The word variable is being used as an insight for training and validation. In the review, each paragraph and record is represented by a neural network with paragraph weight training. Two convolutional layers are involved in the architecture of the proposed model: the sentence layer, which is used to establish a paragraph structure, and the manuscript layer, which is used to convert the sentence vector into an equivalent vector representation. According to Li et al. [6], their suggested system is validated using a dataset including reviews of hotels, restaurants, [6] and other information. The findings of this study demonstrated Cross-domain efficiency as shown in figure 7. CNN outperformed LSTM in the made by mixing and the single domain. Taking inspiration from this, Zhao et al. [16] developed a word order-preserving CNN technique to identify fabricated customer testimonials as shown in table 10.

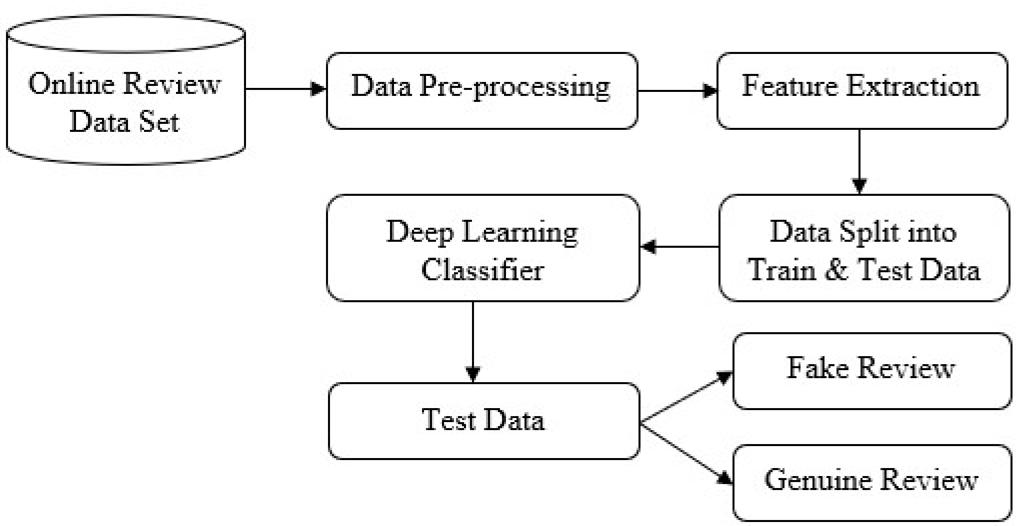

Figure. 7 Architecture of detecting fake online reviews

Table 10. Summary of CNN models in one domain, mix domain, and cross-domain for fake reviews detection.

Ref	Dataset	Methods & Features	Results	Comments
[6]	Hotel, Restaurant, and Doctor.	Sentence Weight Neural Network & Word2vec (Skip-gram).	Accuracy: 79.5% Precision:76.1% Recall:89.9% F1:82.3%	CNN outperformed LSTM in a mixed-domain comparison.
[16]	AMT	Word Order-Preserving CNN & Word2vec and word order.	CNN accuracy:70.02%	CNN cannot handle lengthy articles. So use the hand-annotated method, which is labor-intensive.
[22]	Yelp Chi dataset	Unsupervised neural network model & Word2vec (CBOW). Behavioral Features.	Accuracy: Hotel:65.4%. Restaurant:62%.	Learning review embedding with encodes behavioral and linguistic features is effective.
[37]	AMT dataset Deceptive dataset	DRI-RCNN, Word2vec & Skip-gram.	Accuracy AMT t:82.9%. Misleading:80.8%.	For whatever reason, this model neglected to account for the behavioral aspects that could enhance efficiency.
[44]	Yelp NYC Yelp Zip	CNN & Glove algorithm.	F1-measure:85% for regular reviews and 27% for fake reviews.	They discovered that the quality of the customer's social connections substantially impacted classification accuracy.
[56]	Yelp NYC Yelp Zip	Unsupervised model & Extracted Real behavior features	F1 measure on Hotel:60% & Restaurant :70%.	The importance of link re-weighting in improving performance.
[64]	Hotel reviews from Trip Advisor	LOF algorithm, &Aspect rating & TF-IDF.	Accuracy: 79.6% Precision:79% Recall:80.7% F1-score:79.8.	Aspect rating performed nicely. Fake review identification could be extended by incorporating additional features. Uses a small number of datasets.

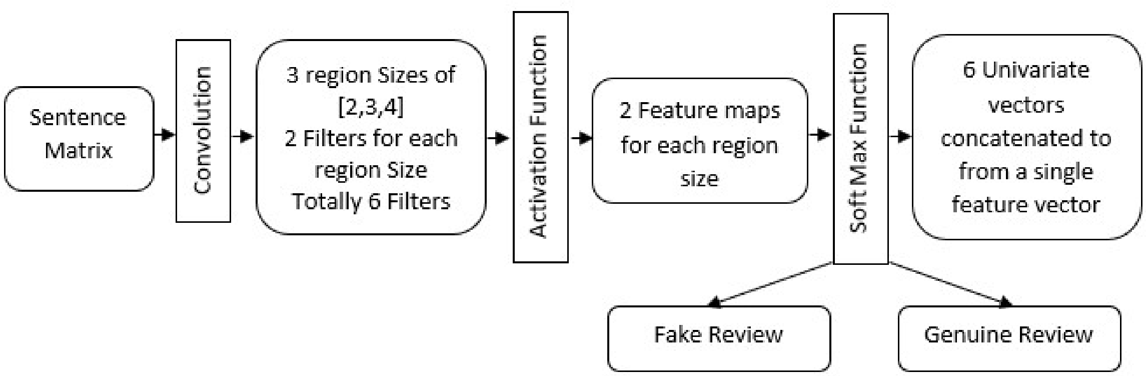

Figure. 8 Architecture of detecting fake online reviews

To produce a vector representation, they utilized the word2vec and the word structure keeping pooling technique instead of the initial max-pooling method, which was more efficient. The feature extracted from the pooling layer has been aggregated in the output nodes, called the concatenation layer, and the AMT dataset [77], those who have analyzed 00 reviews that used the data preparation model presented by Li et al. [84]. Using investigational evidence, it was discovered that the suggested framework beat the current techniques [156, 158] with a precision of 70.02 percent the other hand, CNN proved to be a much more capable method intended for categorizing little messages reviews. CNN required less time for training, whereas RNN was much more effective when dealing with long texts [16] as shown in figure 8. Although the hand-annotated method is effective, it needs many human resources. NN is used to boost the efficiency of every classification technique. By indicating whether a review is behaviorally, grammatically misrepresentative, or both. The ANN technique has developed by Wang et al. [75]. The suggested framework utilized in active to measure training, determined by monitoring behavioral and linguistic trends. The

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

behavior features were extracted using a multi-layer perceptron, and the linguistic features were extracted using a CNN. Therefore, an emotional weight for lingual and behavioral factors was learned using the attention technique. Outcomes from experiments carried out on the yelp data source [8] demonstrated that the suggested process performs well than traditional approaches [8], with a precision of 88.8 percent in the Hotel and a 91% precision in the Restaurant field when compared to the state methods [8], [76]. On the other hand, the proposed technique is essential for improving the accuracy of the classification approach. While some linguistic characteristics were included in the proposed method, behavioral characteristics were included, which meant that the process was insufficient to detect fake reviews. Afterward, et al. [22] developed an unsupervised neural network architecture to handle the cold-start issue (when the latest reviewer publishes an examination view to identify fake reviews based on textual behavioral information of the reviewers. The evaluation text has modeled d CNN, which can capture complicated semantic meaning that is highly hard to convey using conventional features such as the unigram and LIWC [3]. The suggested framework could differentiate between the reviews by integrating text behavioral data. In addition, Trans E is a technique for encoding the framework of a graph, with nodes and edges representing the nodes and edges being utilized to convert based on behavioral [159]. The investigational outcomes on the Yelp dataset [8] demonstrated that the suggested model outperformed SVM in high accuracy, achieving percent in the hotel dataset and preconvention in the restaurant dataset, respectively. [8]. Starting to learn review, on the other hand, is a must. It is much more effective to embed information encoded by behavior to ensure that neither embedding nor feature reduction techniques was utilized to compare the suggested approach. Along with the DRIRCNN detection model for fake reviews, Zhang et al. [37] established an RCNN with word contexts to identify fake reviews in a recent survey. Specifically, the proposed model is composed of 4 layers: A convolutional layer is used to learn the whole vector forward into representing a word. A recurrent neural layer is used to understand the right and left of a phrase's fake and real context vectors. A recursive neural layer is used to learn the right and left for both the false context vectors of the word. AMT and Deception dataset used to test the proposed model The results presented that the recommended model attained the best outcomes with 82.9 percent accuracy on AMT datasets when compared to state-of-the-art techniques such as LIWC and unigram with SVM [77], LIWC attribute, and four n-grams with SVM [140], RCNN [160], trouble arrangement compatibility technique [161], sparsity admixture abstract concept [100], lexical items call for the creation with SVM [123], fully CNN. The suggested model outperformed the state-of-the-art technique on a deceiving dataset by 80.8 percent [77]. However, the proposed framework is time-consuming. Therefore, Li et al. [44] suggested an embedding method to identify fake reviews by impacting consumer reviewer behavior and social interactions. User-item human relationships and consumer characteristics have been combined into a conceivable online review rating expression. The framework suggests items, rating embedding layer upon layer, review embedding systems, and consumer embedding layers. They integrated co-occurrence-based customer behaviors raising be behaviors' performance level for the well-integrated user/item social relations rerandomizes in the network system created by evaluating functions. CNN until utilized W for text encoding. The framework was validated on Yelp NYC and Yelp Zip [72].

The suggested framework outperformed SVM with lingual and app elements [8]. Correspondingly, Li et al. [56] proposed an unsupervised plan to overcome the cold start issue in detecting bogus reviews. Rather than reviewing text for public relationships among consumers, their development has increased link re-weighting to demonstrate behavior patterns. The suggested model has been tested using the [72] Yelp NYC and Zip datasets. The suggested framework scored 60% F1 in the hotel and 70% F1 in the restaurant domain. The proposed framework did not exceed the current process and ignored textual review features to improve the model's performance.

Table 11. Summary of RNN models to detect fake reviews.

Ref & Datasets	Features & Methods	Results	Comments
[3], Gold Standard dataset	Integrated and Discrete features: Word embedding, unigram, POS, LIWC & Bi-GRU with attention	Accuracy on domains: Hotel 81.3% , for Restaurant 87% and doctor 76.3% Cross-domain accuracy: Restaurant 83.7%, doctor 57.3%	Outperform current approaches The proposed model necessitates high
[24], Reviews moblil01.com in Taiwan	Dictionary & LSTM	Accuracy on LSTM:89.4%	The extended LSTM method identified false reviews better than SVM. Compared to the NN approach.
[32],Gold standard dataset	POS and First-Person Pronoun features Glove & Bidirectional LSTM	On cross-domain accuracy, Restaurant 81.3%, Doctor: 66.8% On mix domain accuracy: Hotel 83.9% Restaurant: 85.8%, Doctor: 83.8%	The first-person functionality helps identify bogus reviews. The proposed model necessitates high.
[41], Deceptive Spam Corpus, Four-City, Yelp Zip, Large movie and Drug dataset	Hierarchical CNN GRN deep learning, word2vec, and Multi instant learning methods	Dataset Accuracy Deceptive Spam Corpus: MIL: 90.1%, CNN-GRU:91.9% Four-City: MIL: 82.8%, CNN-GRU: 84.7% Yelp Zip MIL:64.6%, CNN-GRU: 66.4% Big movie. MIL: 87.1%, CNN-GRU: 88.9% Drug: MIL: 78.2%	The conventional CNN and RNN approach only work with small texts. Adding metadata to the suggested model can improve performance.
[50], Gold standard dataset	Wikipedia corpus & Bidirectional LSTM with a self-attention mechanism by word embedding	Accuracy one domain: 85.7%, 84.7%, and 85.5% on Hotel, Doctor, and restaurant, respectively. Mix domain: 83.4% Restaurant: 71.6% and on the Doctor, 60.5%	Fake reviews revealed more emotions than authentic reviews. The model didn't work. Cross-domain outcomes Using adaptation approaches [62, 67] can improve performance.
[70], Spam email. Spam review Political statements	Character-level & Combination of LSTM and CNN	Binary test accuracy of 99.5%	Product inauthenticity can be detected via transfer learning. The model developed employed an n-gram technique.

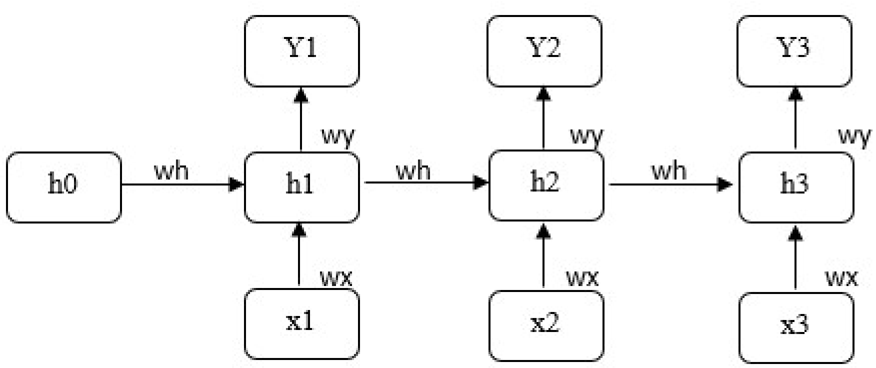

Figure. 9 Architecture of Recurrent Neural Network

Aspect-rating outlier aspect to recognize bogus reviews was recently suggested [64]. They called it extreme rarity detection. Initially, they evaluate the review's element rating using lexicons. For comprehensiveness, the tensor formula was based. Then the studies were classified using the LOF method. Using a TripAdvisor.com dataset, researchers found that the overall element score helped identify fake reviews. Adding reviewer attributes can make it faster.

5.2 Detecting Fake Reviews with Recurrent Neural Network (RNN)

One of the most popular methods customers make purchasing decisions is through customer reviews. The Spiegel Research Centre found that showing customer evaluations increased conversion rates by 270 percent in 2017. Someone who has not tried a product or service writes a fake review. Friends, family members, and even company workers can write them. Bots and corporations that pay people to create false reviews can also generate fake reviews as shown in. This strategy boosts sales or reduces rivals by obtaining fake and favorable competition reviews from other companies. E-commerce and hotel websites allow customers to share their thoughts, feelings, and opinions in the form of reviews. Neuronal networks and Natural Language Processing (NLP) have made recent advances in comprehending people's thoughts, feelings, and behavior patterns more precise. As a result, the reviews might reveal a wide range of emotions, including happiness, rage, astonishment, sarcasm, and disgust. In addition, the app can detect fake reviews, which helps keep the public from being led astray by them. People's emotions may now be extracted more precisely and effectively using Recurrent Neural Networks (RNN). This technology has recently been developed with NLP packages such as NLTK, CoreNLP, and Genism. Learn from the past by using several RNN designs, such as the LSTM and GRU, to generate new outputs. A "hidden" state vector, depending on prior inputs/outputs, influences them in addition to the weights applied to the inputs like a typical neural network.

Consequently, the same information may yield different results depending on the preceding inputs. Because of a high percentage of unidentified lexical items, the hotel and restaurant domain names improved through the doctor. They developed a distinct methodology based on regression models with neurons functionalities from [100] to resolve this challenge. Before the SoftMax layer, those who integrated the neurological and differentiable characteristics. There was 81.3 percent for hotels, 87.5percent for restaurants, and 76.3percenterr doctors. Whenever the classification model received training on hotel reviews, the outcomes were 83.7ppercent of restaurants and 57.3% of doctors. The suggested framework outperformed RNN, CNN, GRNN, and Bi-directional GRNN with one and cross-domain. Adding additional characteristics might also enhance the classifier performance of the model. But the suggested framework is time-consuming. Later, Wang et al. [24] used a lexicon to identify spammers using a recurrent neural network. They created a multilayer perceptron with three levels: an input node (a neuron), an LSTM layer (for feature reduction), and an output layer (a neuron).

The reviewer is either a normal is (0) or a fake is (1). They obtained the stats from Taiwan's moblil01.com and customer reviews websites. They analyzed the info using internal memos. The suggested framework discovered that LSM identified deceptive reviews more accurately than SVM (89.5%). Also, LSTM outperforms RNN in terms of long-term recollection. Other neural network methodologies outperformed the proposed method.

Furthermore, the suggested framework was text-only, ignoring the behavioral and metadata features that could increase implementation. To solve the RNN constraint, Liu et al. [32] developed a BiLSTM standard to understand the document-level descriptions of reviews designed to check false comments depending on combination characteristics. Combining feature interpretation (POS), first-person pronoun characteristics, and content representations (word2vec (Glove)) incorporates features of the suggested method. The AMT dataset [6], which includes three domains, was used to test the system (physician, lodging, and restaurant). The investigational findings demonstrated that the suggested standard beat state-of-the-art techniques like sentence mean, SWNN, SWNNCPOSCI, BiLSTM, and simple CNNCPOSCI. With an 83.9 percent precision in the diverse field, the suggested standard surpassed the state-of-the-art approaches (SWNN, Deep CNN, CNN-LSTM, and CLSTM). Eventually, the findings in one field beat state-of-the-art techniques, with an 83.9 percent efficiency in the hotel field, an 85.8 percent precision in the restaurant domain, and an 83.8 percent precision in the medical field. We can see that the first-person pronouns feature is essential in detecting false reviews based on the model findings. However, the suggested standard necessitates a significant amount of computer power. Jain et al. [41] recently introduced hierarchical CNN-GRN deep learning techniques, and Multi instant learning (MIL) methods are being suggested to accommodate varying reviews in false reviews detection.

Extraction of localized n-gram qualities was accomplished using a three-layer CNN model.

GRN, on the other hand, has been used to discover the semantics among the CNN-extracted characteristics. For example, whether a word's size falls below _fifteen, the last occurrence of the input sentence is deleted across several cases. Datasets like four-city, Yelp [72], Deceptive Spam Corpus [77], Drug Review [164], and Large Movie Review [165] are used to test the proposed model. MIL and CNN-GRN outperformed traditional CNN and RNN in the experiments on these datasets.

On the other hand, the suggested framework works with small paragraphs. An ensemble classifier to identify fake reviews was invented by Zeng et al. [50] in past years. According to their findings, artificial thoughts were more emotional than authentic ones. Synthetic studies often begin with the exact phrases in the first and last paragraphs. The review's middle, beginning, and end will be encoded using four different bidirectional LSTMs in the proposed model. Fake reviews are detected using a convolution of four depictions.

Three different representations were combined into one using the self-attention technique. This technique has been used to combine the two characterizations into a single final image. With a precision of 85.7 percent, 84.7 percent, and 85.5 percent, the suggested framework outperformed other techniques such as SWNN and SAGA in a (Hotel/Doctor/Restaurant) as a single domain on the AMT dataset. In addition, in the combined field, there is. It had an overall accuracy of 83.4%. It was suggested that the existing model did not perform well cross-domain, with 71.6% in the restaurant industry and 60.5% in the medical field. Dhamani et al. [70] initiated NN and transfer learning to combat social media disinformation. They developed an ensemble approach using LSTM and a sentence CNN. The suggested technique could move to another area using class labels with one field. Trained to identify goods that are not genuine can also be done using transfer learning, which has shown promise. On the other hand, N-grams were used in the proposed system because it was more straightforward.

5.3 Detecting Fake Reviews with Generative Adversarial Network (GAN)

GAN has proven highly efficient in various fields, including image processing [166]. There are two main principles in GAN's framework. One generator generates reviews, and the other is a discriminator that calculates whether an online review is authentic or fake. In Fig. 5, we show a straightforward design for the GAN procedure to illustrate its simplicity. First and foremost, the generator generates an information example and a discriminator that classifies the information as either genuine (training) or false (made by a generator). The generator aims to create a few samples similar to the essential information to deceive the discriminator. Therefore, the discriminator must accurately differentiate between different data samples. In this part, we will summarise the GAN techniques that are currently available in the publications, as demonstrated in Table 10. The fake Generative Adversarial Network (FakeGAN) example suggested by Aghakhani et al. [7] was designed to deal with the scarcity of artificial review detecting datasets due to the model's semi-supervised nature. One generative and one discrimination standard must also be used [167]. To fix the generator's intersection problem with two suggested discriminators and make a greatly more powerful generator: A distinction is produced between fake and genuine reviews in these categories.

The second one distinguishes samples taken from a dispersion of fake reviews and reviews made by the LSTM productive classifier. To train the generator on counterfeit checks, maximum likelihood estimation is employed.

Table 12. Summary of GAN models to detect fake reviews.

Dataset & Ref	Features & Method	Comments
AMT, [7]	Glove2vec & Fake GAN & Accuracy:89.2%	Didn't outperform modern approaches. GAN is not suitable for text classification due to its instability, making hyper-tuning difficult. This suggests improved hyper tuning conditional GAN.
Hotel, Restaurant domain Yelp Chi, [14]	Word2vec (CBOW) behavioral Features & GAN & Hotel 80% Restaurant 75.6%	More data can be utilized to enhance the review embedding representation.
[25] Hotel, Restaurant & Yelp Chi	Actual behavior features & Behavior features are generative and adversarial networks. Hotel 83% Restaurant 75.7%	The suggested model didn't work in restaurants. This implies constructing a model for each part.

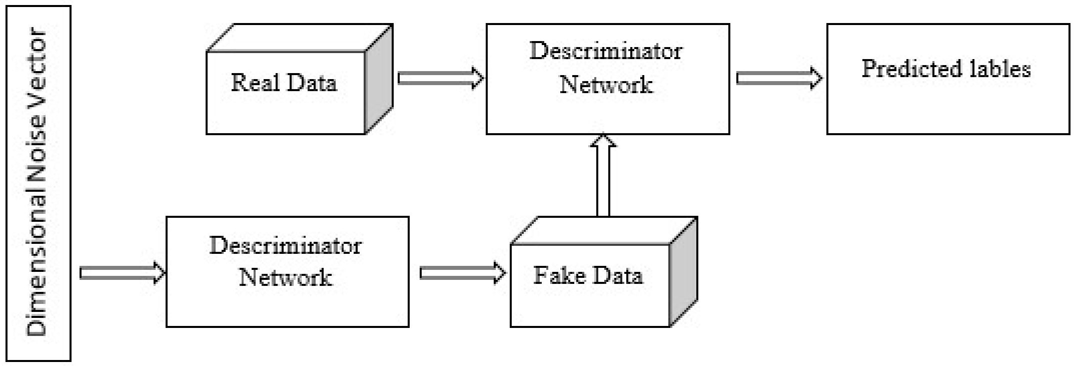

Figure. 10 Architecture of Recurrent Neural Network

According to the outcomes, the suggested framework obtained a precision of 89.2 percent on the AMT dataset [77]. However, the proposed framework did not perform the current technique. Moreover, the results indicate that, due to the strength of GAN, it is not sufficient for reader categorization. Furthermore, hyper-tuning is difficult for the suggested framework to accomplish. Moreover, You et al. [14] used deep learning methods to incorporate essential features from diverse aspects to deal with the cold opening issue in detecting a fake review. They suggested a system that objects, reviewers, and studies in binary code with properties like date, price range, and location. Furthermore, they offered a domain classifier to help transfer knowledge from one field to another. The first layer included many criteria in the model. The second layer caught (entity-entity), (entity-attribute), and (attribute-attribute). The third layer implemented a domain classifier. The method proposed included three layers. The suggested model was analyzed using the Yelp Chi dataset generated by Mukherjee et al. [8] and is available online. The proposed framework outperformed the SVM in precision, achieving 80 percent in the hotel domain and 75.6 percent in the restaurant domain. When dealing with complex start difficulties, the generative adversarial network is invaluable.

On the other hand, the proposed methodology wasn't evaluated compared to other embedding techniques. Tang and colleagues later addressed the cold start difficulty in identifying fake reviews [25] and suggested a generative adversarial network approach to deal with the issue. For new subscribers who do not have any characteristics, synthesized behavior characteristics are obtained for them.

First, six meaningful features for new and frequent clients were gained. Because they employed efficiently accessible features as input, the attributes are created utilizing a GAN method. GAN's generator comprises six layers, which are as follows: The first three layers have been used for normalization reasons and to obtain characteristics that have been easily accessible. Then we used the other three layers to synthesize behavior using readily available features. A new client is then used to train the generator using GAN's discriminator. The suggested framework was validated on the Yelp Chi dataset [8], which contains two distinct domains] and was found to be effective (Hotel and Restaurant). Compared to the existing methods [22] and [14], the suggested framework outperformed them in precision, achieving an accuracy of 83 percent in the hotel dataset and 75.7 percent in the restaurant dataset. Furthermore, the achievement of the classifier was enhanced when multiple characteristics were merged. On the other hand, the suggested framework was unsuccessful in identifying fake reviews across domains.

5.4 Other Neural Network Methods

This section will summarize the other published neural network models (Table 11). Rather than depending on specialist expertise, Wang et al. [76] built a novel spam detection prototype based on relationships between reviewers and products. A three-mode tensor was created from the relationships between two things. Next, a tensor factorization method known as RASCAL [168] was used to accurately classify the good or service vector representations and reviewers. An SVM classification model has then injected the final iteratively interpretation of the review, which

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

is the final step. It must have been decided to use the dataset of Yelp Chi [8] to verify the suggested framework. In the hotel domain, the results demonstrated that the proposed standard exceeded the state-of-the-art approach [8], [72], with a precision of 85.9 and 87.8 percent, respectively. The suggested framework demonstrated that the relationships between the product and reviewer are essential to enhancing the accuracy of the categorization standard. Wang and colleagues [17] developed a time series multi-dimensional model for detecting only one fake review in the wild. A new index has been developed to measure the authenticity of reviewers by combining their reliability and expert knowledge into a single score. The ranking methodology was designed to summarize all spam across a range of areas to detect anomalous elements of time series. A bogus review in a time series decreases the window size rate by one unit. The researchers found that their proposed framework outperformed the mean RHR in human evaluation in their study. According to the investigation, numerous studies were decided to post simultaneously on various days between 2009 and 2010.

On the other hand, the suggested framework does not use accuracy, F1 measure, recall, precision, and to assess the technique's efficiency. A pattern recognition framework was developed by Heydari and colleagues [5] to recognize mistrust that appeared during suspect time frames based on metadata and rankings deviation characteristics. It is possible to install a time series to identify oscillations in various reviews for every good or service. It is necessary to utilize a sliding window to encapsulate the suspect durations and identify trends. According to the results obtained from accurate Amazon.com data sources, the proposed method was conducted well in identifying fake reviews, according to the results obtained from accurate Amazon.com data sources, which had an F-measure of 86 percent. The advent of the suggested protocol was supposed to focus on unreliable intervals rather than decrease the number of costly equations in the counting phase. Aside from these issues, combining metadata, such as an IP address, could increase the suggested scheme's effectiveness. Later, Li et al. [27] discovered phishing reviews on Amazon using a paragraph graded neural network model (SWNN).

Almost any sentence is connected to the weight over the proposed method, then transforms the sentence into a document vector. A paragraph was made up of words from different reviewers. Then they upgraded POS and First-Person Pronoun to verify the review. They were tested on the AMT dataset [6] hotel, restaurant, and doctor domains. The unigram characteristic had the best results in the restaurant field, with 78.5 percent accuracy. In the physician field, merged features delivered the best results, with just a precision of 61.5 percent SWNN surpassing state-of-the-art techniques [119] [169] made by mixing domains, with an accuracy of 80.1 percent F1 score is being used as a measure o one. The following results have been obtained: 83.7 % hotel, 87.6 % restaurant, and 82.9 % doctor. Moreover, in combination and cross-domains, we cannot forecast accurate results. Noekhah et al. [45] identify a single false review, a group of reviewers, and just one reviewer at a time by an unsupervised Multi-Iteration Network Structure. Inter-relationships (associations among reviewers) and interpersonal and inter (associations among good or service, reviewers, and reviews) are being used as extracting features in the new proposal. The suggested framework accomplished a 98 percent precision with features, a 74 percent precision based on behavioral characteristics, and a 69 percent precision with internal structures on the Amazon.com set of data. However, who did not try comparing it to other techniques to demonstrate the performance of the proposed scheme. They didn't employ all metadata attributes to help the classifier. Yuan et al. [57] recently suggested a new fusion focus network for detecting fake reviews by computational methods at the good or service and user levels. A consumer needs multicount attention if presented to obtain factors potentially from paragraph interpretation. The user-product understanding was then learned using fusion attention elements and orthogonal decomposing. Finally, they defined feedback as consumer-product interactions. To encapsulate the product-review-user association, they have been using TransH, a prototype for embedding a knowledge graph in vectors [170]. The model was tested using the Mobile01 Review [171] and Yelp [72] datasets.

The outcomes show that the suggested framework outperforms current techniques like SVM with content and behavioral data. [97], [171], RSD [150], page [72], TDSD [76], Couple Hidden Markov Model (CHMM) [81], Spam2Vec [172], CNN-GRNN [3], SWNN [27], ABNN [75], AEDA [14]. The Mobile-first post dataset received an 86.96 percent F1 score, while the Mobile Reply dataset received a 48.37 ppercentF1 score. It also received an 83.24 percent UC on the Yelp Chi dataset, an 84.78 percent on the Yelp NYC dataset, and an 87.28 percent UC on the Yelp Zip dataset. The suggested framework demonstrated that both the item and consumer levels are essential in identifying fake reviews. Cao et al. [65] provided a model for detecting false reviews that implicitly combines fine-grained and coarse characteristics to extract semantics from reviews. A coarse-grained combination of two NN layers and Latent Dirichlet Allocation techniques was used (LDA). In addition, DL methods have been utilized to learn the fine-grained characteristics concurrently with the coarse-grained features. Finally, an SVM algorithm is trained to determine whether the review is legitimate or fake. The outcomes obtained from a gold standard data source [100] and a real-life dataset [8] demonstrated that the suggested framework might enhance the accuracy in the one-domain and the combination scenarios. Furthermore, LDA combined with Content CNN made the highest outcomes on the datasets in the one-domain and the united. Again, utilizing coarse-grained characteristics outperformed using fine-grained attributes in terms of overall efficiency. Moreover, compared to the simple model, the suggested framework has a higher level of time difficulty. Similar hybrid deep learning technique [69] to gain review semantics to detect fake reviews. The proposed methodology was divided into three stages: first, they extracted the review encoding using two NN architectures (the Paragraph Vector Distributed Bag - Of - words and the Denoising Autoencoders). Then, they used the review incorporating to train the neural network system. In the feature extraction tokenization, two methods are integrated and fed to a fully connected layer, determining whether the review is fake. Using a dataset of gold standard [77], the suggested framework significantly performs the techniques with 92.5 percent, superior to existing methods. Contributing advanced capabilities, such as an emotional component, might, on the other hand, boost the effectiveness. According to their findings, Guo et al. [71] suggested a graph NN framework to detect spammers by simultaneously incorporating occasional connections and stable interactions in the network. Extraction of the rare interactions has been accomplished using the parameterized technique [173], while the mode modeling-lasting relationship has been performed using a direct vector representation encoding technique. DL graphs were created to design the characteristics of conversation. Present research work from two different data sets, it was discovered that the suggested framework surpassed standard strategies like CNN, MLP, SVM, and LSTM in terms of accuracy.

In machine learning, neural networks have some of the most successful methods. Deep understanding has been used in this topic of investigation, and meaningful results have been obtained. Deep learning does not require the features to be extracted from input data; instead, these could be typically extracted from the input dataset without any need for previous knowledge gained or interference. As previously stated, when used to identify fake reviews, such approaches have some real restrictions. One of the most significant issues affiliated with DL techniques is that they do not provide extensive insight into how people learn. DL can be judged as a "black box" concept for this discussion because it lacks facts and information to describe the results. A further disadvantage of deep understanding is that it demands more data than conventional ML, which indicates we could use supervised learning models with limited data. Moreover, deep learning techniques necessitate a significant investment of computation power.

6 EXPERIMENTS

Several strategies have been presented over the past few decades to address a wide range of issues in the online platform (false online reviews, ratings, fake news, disinformation, anomaly detection, etc.). Identifying and solving gaps in research in many fields is an ongoing challenge for researchers. Deepome a prevalent approach in recent years, in recent years

Here, a first-hand estimation of the accomplishments of 7 good deep learning techniques on different datasets is described. The results are discussed in detail in the following section. LSTM, HAN, and convolutional -LSTM methods at the character level One can choose from many ways, such as the Distil BERT, HAN, BERT, and Roberta. The primary objective is to determine whether such methods can detect fabricated reviews. Some of these techniques are being used by professionals in different fields [174] 179], including finance and medicine.

Moreover, they have still not been applied in fake review identification. Nevertheless, the effectiveness of such techniques in identifying bogus reviews was presented in the study, which can serve as a base point for further investigation. We used a dataset for the early developmental stages in this research, one for each group. The first data file is the "Yelp Consumer Device dataset" [79], which clambered via review datasets related to

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

web procedures from Yelp.com, and the data set is the "Yelp Consumer Electronic dataset" [80]. They assigned labels to them depending on the content and customer data characteristics. The rule-based technique has been used to annotate the information in this dataset. Include these examples: the data set has been built based on specific regulations, which determined that a review had been fraudulent when distinct people share thoughts of the same/different good or service. The above dataset offers essential information to help researchers construct fake review prediction models that may be utilized in the real world. Finally, the "deception dataset"[100] was generated using TripAdvisor and Amazon Mechanical Turk websites in Chicago and included 3,032 reviews from various sectors (Hotel, Restaurant, and Doctor). This form of data has been widely used in papers [3, 4, 12, 27, 29, 32, 37, 65]. We've aggregated the reviews from the three categories to make things easier. Each domain (i.e., a multi-domain detection technique) will be left to future researchers to complete. As the preceding section demonstrates, the following steps must be followed first in the fake review detection algorithm.

6.3 Dataset Pre-Processing

Data points have been pre-processed throughout this phase to remove noise, including stop words, URLs, emojis, and other ambiguous characters. The pre-processing was decided to carry out using the NLTK toolkit1. An open-source repository is applied in the field. First, we divided the content into tokens by tokenization. After that, we eliminated the stop-word in the text classification stage, which caused the noise. Ultimately, we use the steaming technique to break down the words into their constituent parts. Reviews in the duplicity data and the consumers of electronic in yelp datasets are shown in Table 12. Finally, we merged the three-domain reviews into a single deception dataset for simplicity.

7 FEATURE EXTRACTION

Getting the most exciting and precise source data is critical to enhancing performance and results. The extraction of features is one of the essential parts of this process. GloVe embedding techniques with 100 facets [116] were used to create neural network models from data that had already been pre-trained. To build word representations, the GloVe uses an unsupervised learning algorithm training on a small amount of data: one billion phrases. It has demonstrated impressive outcomes in the identification of fake reviews. In the vector space, a straightforward algorithm is being used to impose word vectors to acquire sub-linear connections GloVe. It outperforms Word2vec in the word metaphor activities, because of which it is recommended. Furthermore, by focusing on the relationships among similar words rather than single words, Glove provides a more practical interpretation of word vectors. The Glove also provides good weight to extremely frequently used word pairs, allowing meaningless stop words such as "the" and "an" not significantly impact the overall process groups C. ALGORITHMS The neural network techniques and the transformers that we employed in our tests are described in this section.

C-LSTM is a trademark of C-LSTM. The C-LSTM uses CNN to retrieve a sequence of relatively high phrase descriptions. It inputs it into an LSTM to produce the sentence statement [174]. The convolutional layer performs a matrix-vector expression to every lexical item. The LSTM algorithm continues to spread past data throughout the neural network hierarchy. In our research, a CNN is developed to learn higher representations of n-grams on a pre-trained word vector. Then, to learn sequential correlations using sequence similarity models, the feature mappings CNN are created as sequential windows characteristics to act as the source of LSTM. This transforms every word into an n-gram characteristic that activates factors in phrases. A single LSTM layer and a single convolutional layer with 128 filters were both used in our research. The data was then input into an Lstm model with a loss of 0.2 and output parameters of 100. Lastly, we classified the review as fake or genuine using the sigmoid activation function in the output nodes.

7.3 Character Level C-LSTM

This algorithm [175] identifies as inputs a series of coded words that has been encoded. Encoding is accomplished by providing a predetermined amount of alphabet letters for the input sentence, then analyzing each letter, including one encoding to produce the desired result. The letters are then transformed into vectors with a specified size. Due to the reversal of the quantifying ordering for each line, the last letter read appears near the start of each result, making it simple to correlate weights with the most recent measurement for entirely fully connected. Researchers used letters from the review data to create a character-level embedding layer in our research when collecting letters from the review data. Convolutional filters 3 and 5 were applied

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

after a layer of convolution units and before a final layer of convolution units. We used two maximum pooling and a dropout of 0.2 in our experiment. Next, we developed a bi-LSTM with fully associated layers. A Re, LU mechanism called the ReLWeee employed the sigmoid activation function for the node's output in conjunction with optimization on ADAM.

7.4 Hierarchal Attention Network (HAN)

HAN stands for Hierarchical Attention Network. However, it is a method that has been developed for capturing the entire data model. (Word encoder, word attention, sentence encoder, and sentence attention) are implemented using Bidirectional GRU in the HAN model, as per [176]. First, the total size in our experiment was set to 200 characters, and then a Bi-GRU with 100 outputs parameters was supplied to the attention layer. To construct a phrase encoder time scattered layer, we used the input from the word encoder as input. We used a 0.001 ADAM learning rate to optimize our model and get the best outcomes for our final step.

7.5 Convolutional Han

We used a 1-d convolution layer before every two-way GRU layer in this model to capture high-level input properties. The attention layer receives the word from this layer, which reviews and feeds it. The Bi-GRU attention layer received 100 output dimensions by the HAN structure, as did the rest of the system. We used an ADAM optimizer with a 0.001 learning rate to fine-tune our models.

7.6 Bidirectional Encoder Representations from Transformers (BERT)

BERT is a converter and classifier that has already been pre-trained [177]. BERT has been pre-trained on more than 800 million words of literature and more than 2500 million English Wiki entries. BERT scans the entire string of dishes at once, allowing the model to understand the word's context from its surroundings instead of reading from right to left or left to right (right and left of the word). Twelve converter layers and 768 hidden layers comprise our BERT model, utilizing a layered architecture. One word at a time, the version progressed.

The input words are tokenized and translated to BERT input IDs. Every phrase had a Classifier Token and a SEP (separated Segment Token).

The input mask 0 indicates padding integers and one unpadded bit. Each converter received the token encoding list and produced a feature representation of the same length. The CLS outputs on the 12th converter layer comprising predictions of likelihood vector changes have been classified.

7.7 DistilBERT

BERT's computation time, constant input distance size, and phrase piece embedding issue have been addressed by DistilBERT [178]. Dilbert seems to have the same paradigm as BERT, but with fewer layers, tokens kind encoding, and no pooler. Our DistillBERT design had six layers of layered transformers with 12 self-attention layers and 768 hidden levels. First, we segmented the given words and input IDS. Our DistilBERT classifier was then fed the padding source IDs for classification tasks.

1) Robustly Optimized Bert Approach (RoBERT)

By training the classifier lengthier, training on lengthier sequencing, and omitting the predictions for another phrase, Roberta could outperform the BERT transformer's effectiveness [179]. Roberta is also pre-trained on 63 million news pieces from the Commons Crawl News databases, English Wikipedia, and novels. In this study, the Roberta tokenizer has been utilized for encoding the input to the token and classifying them as inputs ids. The duration of such IDs has been extended with a fixed value to eliminate any potential for sequential variations. The tokens were then evaluated to understand the features of the categorization of sentence pairs.

8 RESULTS AND DISCUSSION

Deep learning methods and converter structures are specifically described in this section, as is the effectiveness of deep learning techniques. We employed the identical characteristics initially recommended for the architecture to conduct such demonstrations. In addition, we separated every dataset into three groups: training, validity, and testing to complete tests.

Based on such relevant constraints, the algorithm's efficiency in recognizing false reviews in terms of accuracy, precision, recall, and F1 score, is shown in Table 13.

When contrasted to peer techniques, Roberta outperformed them across both datasets, achieving 70.2 percent for correctness, 65.2 percent for accuracy, 61.5 percent for recall, and 61.5 percent for F1-score comparison to 61.5 percent for precision. The deception dataset attained an average accuracy of 91.02 percent, precision rates of 92.5 percent, 90 percent and F1 scores of 90.5 percent, ana decisions of 90 percent, recall rates of 90 percent core, spending. Oddly enough, their effectiveness on the deceptive dataset is much better than that of the Yelp database. It is challenging to recognize bogus reviews on the Yelp dataset (70.2 percent correctness) if only a dataset is valid and fake review data overlaps. As a result, this is a particularly demanding problem. On the other hand, deceptive data is a representation of semi-real information. In addition to BERT, other transformers' approaches also made impressive progress on both data in this study. This concludes that transformers algorithms were significantly more effective at identifying negative reviews than other techniques Because they've been trained on many datasets.

Using these techniques and creating new ones might be a beneficial foundation for future use and the development of new sorts in identifying fake reviews. The convolutional neural network (HAN), the convolutional neural network (CONV), and the char-level C-LSTM have all had issues with performance. In two ways, this is understandable: For starters, to learn and operate well, such techniques take a significant quantity of data. Both datasets contained a few thousand comments, which may not have been enough to distinguish between authentic and bogus testimonials. Two reasons for this are that such algorithms require a lengthy tuning process to achieve the best outcomes. When we conducted our studies, we employed predetermined variables from such models that have been published in the literature that might not have been acceptable for using counterfeit products fake. Furthermore, this report offers an in-depth analysis for enhancing the algorithm's efficiency in the coming to increase the efficiency of the false approach noneffective.

9 PERFORMANCE METRICS.

We utilized a variety of indicators to evaluate how well algorithms performed. Most of them have been primarily dependent on the confusion matrix. We may create a confusion matrix using the four classification model parameters: true positive, false positive, true negative, true negative, and false negative.

9.3 Accuracy.

When it comes to metrics, accuracy is frequently the most employed. It represents the percentage of accurately predicted observations that were either true or incorrect. It is possible to measure the accuracy of a model's performance by using the following equation:

$$\text{Accuracy} = \frac{(TP+TN)}{(TP+FP+FN+TN)} \quad (1)$$

9.4 Recall.

The recall represents the total number of positive classifies made from the genuine class. Our example indicates the proportion of reviews projected to be true out of the total number of reviews forecast to be true.

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

9.5 Precision:

Precision score, on the other hand, is the percentage of true positives to all projected actual occurrences. Precision is the number of reviews that have been correctly anticipated (actual) out of all the reviews that have been accurately forecasted (predicted).

$$\text{Precision} = \frac{TP}{TP+FP} \quad (3)$$

9.6 F1 Measure:

The F1 score reflects the trade-off between precision and recall in each situation. It computes the harmonic mean of the difference between the two values. It considers both the false positive and false negative findings, which is beneficial. The F1 score could be determined with the help of the equation below.

$$\text{F1} - \text{Measure} = \frac{2 * \text{Recall} * \text{Precision}}{\text{Recall} * \text{Precision}} \quad (4)$$

10 CHALLENGES IN DETECTING FAKE REVIEWS

In recent decades, a great deal of effort has been put into increasing the validity of digital content. However, the fact remains that, notwithstanding the improvement that has been done, there are still areas that need improvement. This part will discuss the gaps in this study area and open avenues for further investigation.

10.3Group of spammers detection.

According to the research, finding out who the spammers are is essential for spotting fraudulent reviews (Mukherjee et al., 2012). Many spammers result in the dissemination of bogus reviews at specific real-time intervals. As a result, researchers determined it by considering research that concentrated on bursting trends to identify fake reviews. As a result, researchers were able to identify counterfeit studies with fantastic precision. The exploration of bursting trends utilizing innovative technology to detect spammers is an area that needs further examination.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

10.4 Explicable Fake Review Detection Model.

Natural language processing benefited greatly from deep learning, with particularly impressive outcomes. However, it is classified as a "Black Box" since it does not possess descriptive information that may be used to provide additional details of the results. Moreover, deep learning algorithms for detecting fake reviews are unclear and cannot be understood. As a result, it is hard to place confidence in the model's outcomes. In other words. For example, how have some deep learning techniques outperformed other models on one database while outperforming different algorithms on another data? So, what exactly do deep learning techniques learn? The investigation of readability could be carried out depending on basic ideas. However, as of now, nothing research has been done to explain the bogus review detecting model. This necessitates the development of easy-to-understand counterfeit review detection technologies.

10.5Handling Concept drift problem.

Earlier techniques for identifying false reviews across real-world applications where the attributes of the studies vary over time due to the rapidly changing reviews [51] may not always be suitable for use in situations where the reviews' characteristics change with time. A further consideration is that, in application scenarios, the prediction must be modified regularly [51]. As a result, there is a requirement for a practical approach that could deal with the idea drift issue in real-world situations.

5.3 One class classification model.

Using a class classification technique in a real-world application could remedy large datasets. Examples include using a class condition called Randoms field to the Twitter dataset for anomalous data analyses [181]. However, unstructured real-world data can be handled using one-class classification methods like OSVM [182] and Non-OSVM frameworks [183] [185]. Hence the need to investigate unlabeled false review datasets to fix this issue of a shortage of datasets in identifying fake reviews.

5.4 Cross-domain fake review detection.

Communication across different domains must be managed correctly. However, when it comes to counterfeit review identification, the problem of a shortage of annotated resources is a disappointment. An important research field is applying a strategy that has been trained and verified in the input space to the target task. Because the current literature concentrated solely on one area of fake review identification, the presented approaches were ineffective when trained in one place and then evaluated in another. The researchers [80] tested the model in one domain and assessed it in another. The efficiency has been significantly reduced when comparing different results to the findings in almost the same domain. Cross-domain bogus evaluations require deeper investigation and research [67], [186].

10.6Multilingual fake reviews detection.

The detection of bogus reviews is also incorporated into the interdisciplinary assessment process. Consumers could post comments in various accents, including English, Chinese, Malay, and Arabic, depending on their preference. Users can even write a review in their native language. Yet, only a few studies [97] and [187] have effectively used fictional review data from various dialects. Since spammers publish quickly and

then take effect from those other sites, it isn't easy to spot them. A language translation to convert the English review into every language the spammer demands appears to be an alternative solution.

11. CONCLUSION

This paper provided an exhaustive study of the most important documents on machine learning-based fake review identification that have been published to date. First and foremost, we examined the various extracting features algorithms that multiple scholars have developed. After that, we went over the existing datasets and the methodologies used to create them. We discussed several standard machine learning techniques and neural network models used for fake review identification and presented them in tabular format. Improved extraction of features and classifiers design are two ways in which conventional statistical machine learning can boost the effectiveness of text categorization models. On the other hand, deep understanding enhances performance by improving the presenting learning approach, the algorithm's architecture, and extra knowledge. A comprehensive examination of several neural network model-based machine learning and converters that have not been employed to identify false reviews was also presented. The results revealed that Roberta had the Best accuracy on both datasets, demonstrating its, superiority.

Table 13: Fake online reviews: Literature review, synthesis, and directions for future research

Journal of Theoretical and Applied Information Technology
15th July 2022. Vol.100. No 13
© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

Datasets (Name & Sources)	Data used by	Descriptions	Rating	Text	Image
Amazon [184]	[81, 92,98, 119,121,145, 185]	reviews are 5,838,041, reviewers are 2,146,057 and products are1,230,915.	No	Yes	No
Amazon [192]	[83]	Kitchen, Books, DVDs, and Electronics are the four product kinds reviewed.	Yes	Yes	No
Amazon [193], Amazon [194], Amazon [195]	[158, 186]	142.8 million reviews of various products have been posted online.	Yes	Yes	Yes
TripAdvisor [187]	[12, 26, 72, 75, 76, 78, 85, 91, 93, 96, 101, 102, 105, 107, 110, 113, 131, 140, 146, 148, 150]	There are 800 positive reviews, 400 of which are genuine and 400 of which are false.	No	Yes	No
TripAdvisor [196]	[12, 26, 55, 72, 75, 76, 78, 85, 93, 96, 102, 107, 110, 148]	a total of 800 unfavorable reviews, 400 of which are legitimate and 400 of which are fraudulent.	No	Yes	No
TripAdvisor and Yelp [182]	[88, 94, 105, 146, 149]	One thousand two hundred genuine reviews and 1636 fraudulent reviews from three categories: hotel, restaurant, and physician.	No	Yes	No
Yelp [197], Yelp [198]	[12, 100, 103, 109, 121]	Reviewers have leftover 67,000 reviews for over 200 hotels and restaurants. 160,225 users have written 359,052 reviews for 923 restaurants.	Yes	Yes	No
Yelp ,[198]	[24,103, 121]	260,000 people left 608,598 reviews of 5,044 eateries.	Yes	Yes	No
Yelp [189]	[24, 29, 106, 117, 121, 123,148]	9456 real reviews and 9456 false reviews from four cities in the United States.	Yes	Yes	No
Dianping [125]	[125]	9067 reviewers provided a total of 9765 ratings for 500 different eateries.	No	Yes	No
Hotel Review [51]	[51],[54], [55], [56], [59], [79]	Context: internal reference price & willingness to pay, Data: 766 responses	Yes	Yes	No
Restaurant reviews [81]	[81], [60], [82], [83], [158], [97], [84], [76], [77], [62], [63]	The effect of review ratings on usefulness and enjoyment, 5090 reviews of 45 restaurants.	Yes	Yes	No

Furthermore, Roberta's recall, precision, and F1 score showed the system's effectiveness in recognizing bogus reviews. Lastly, we outlined the existing gaps in this study area and the prospective future directions for achieving robust results in this domain in this section. Based on the current research, most work has concentrated on supervised machine learning to identify false reviews. While supervised machine learning could be used to forecast whether a review is incorrect or not, it requires a tagged dataset, which could be challenging to come by in the field of study spam detection. We discovered that the most frequently utilized dataset in the current research was created via a wiki approach, which we attribute to difficulties of getting tagged datasets. Because these data do not represent the fake review in practical applications, evaluating the machine learning approach on such datasets is not recommended. Because of this, it is preferable to evaluate the classifier on practical uses, as this will assist us in statistical modeling that is efficient in the real world. We feel that academics who have a thorough understanding of the critical components of this discipline will find this study to be quite helpful. It provides an overview of the most significant developments and highlights the anticipated future trajectories.

REFERENCES:

[1] F. Cardoso, R. M. Silva, and T. A. Almeida, ``Towards automatic _ltering of fake reviews,'' Neurocomputing, vol. 309, pp. 106_116, Oct. 2018.

[2] L. Da Xu, W. He, and S. Li, ``Internet of Things in industries: A survey,'' IEEE Trans. Ind. Information., vol. 10, no. 4, pp. 2233_2243, Nov. 2014.

[3] Y. Ren and Y. Zhang, ``Deceptive opinion spam detection using neural network,'' in Proc. 26th Int. Conf. Comput. Linguistics: Tech. Papers (COLING), 2016, pp. 140_150.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[4] N. Jindal and B. Liu, ``Opinion spam and analysis,'' in Proc. Int. Conf. Web Search Web Data Mining (WSDM), 2008, pp. 219_230.

[5] A. Heydari, M. Tavakoli, and N. Salim, ``Detection of fake opinions using time series,'' Expert Syst. Appl., vol. 58, pp. 83_92, Oct. 2016.

[6] L. Li, W. Ren, B. Qin, and T. Liu, ``Learning document representation for deceptive opinion spam detection,'' in Chinese Computational Linguistics and Natural Language Processing Based on Naturally Annotated Big Data. Nanjing, China: Springer, 2015, pp. 393_404.

[7] H. Aghakhani, A. Machinery, S. Nilizadeh, C. Kruegel, and G. Vigna, ``Detecting deceptive reviews using generative adversarial networks,'' Proc. IEEE Secure. Privacy Workshops (SPW), May 2018, pp. 89_95.

[8] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, ``Fake review detection: Classification and analysis of real and pseudo reviews,'' Univ. Illinois Chicago, Chicago, IL, USA, Tech. Rep. UIC-CS-03-2013, 2013.

[9] R. Yafeng, J. Donghong, Z. Hongbin, and Y. Lan, ``Deceptive reviews detection based on positive and unlabeled learning,'' J. Comput. Res. Develop., vol—52, no. 3, p. 639, 2015.

[10] R. Y. K. Lau, S. Y. Liao, R. C.-W. Kwok, K. Xu, Y. Xia, and Y. Li, ``Text mining and probabilistic language modeling for online review spam detection,'' ACM Trans. Manage. Inf. Syst., vol. 2, no. 4, pp. 1_30, Dec. 2011.

[11] A. Heydari, M. A. Tavakoli, N. Salim, and Z. Heydari, ``Detection of review spam: A survey,'' Expert Syst. Appl., vol. 42, no. 7, pp. 3634_3642, May 2015.

[12] Y. Lin, T. Zhu, X. Wang, J. Zhang, and A. Zhou, ``Towards online review spam detection,'' in Proc. 23rd Int. Conf. World Wide Web (WWW Companion), 2014, pp. 341_342.

[13] Z. Hai, P. Zhao, P. Cheng, P. Yang, X.-L. Li, and G. Li, ``Deceptive review spam detection via exploiting task relatedness and unlabeled data,'' in Proc. Conf. Empirical Methods Natural Lang. Process., 2016, pp. 1817_1826.

[14] Z. You, T. Qian, and B. Liu, ``An attribute enhanced adaptive domain model for cold-start spam review detection,'' in Proc. 27th Int. Conf. Comput. Linguistics, 2018, pp. 1884_1895.

[15] Z. Sedighi, H. Ebrahimpour-Komleh, and A. Bagheri, ``RLOSD: Representation learning-based opinion spam detection,'' in Proc. 3rd Iranian Conf. Intell. Syst. Signal Process. (ICSPIS), Dec. 2017, pp. 74_80.

[16] S. Zhao, Z. Xu, L. Liu, and M. Guo, ``Towards accurate deceptive opinion spam detection based on word order-preserving CNN,'' 2017, arXiv:1711.09181. [Online]. Available: http://arxiv.org/abs/1711.09181

[17] Y. Wang, S. C. F. Chan, H. V. Leong, G. Ngai, and N. Au, ``Multi dimension reviewer credibility quanti_cation across diverse travel communities,'' Knowl. Inf. Syst., vol. 49, no. 3, pp. 1071_1096, Dec. 2016.

[18] F. Khurshid, Y. Zhu, Z. Xu, M. Ahmad, and M. Ahmad, ``Enactment of ensemble learning for review spam detection on selected features,'' Int. J. Comput. Intell. Syst., vol. 12, no. 1, pp. 387_394, 2018.

[19] B. Pang and L. Lee, ``Opinion mining and sentiment analysis,'' Comput. Linguist, vol. 35, pp. 311_312, Mar. 2009.

[20] E. Fitzpatrick, J. Bachenko, and T. Fornaciari, ``Automatic detection of verbal deception,'' Synth. Lectures Hum. Lang. Technol., vol. 8, no. 3, pp. 1_119, Sep. 2015.

[21] B. Liu, ``Sentiment analysis and opinion mining,'' Synthesis Lectures Hum. Lang. Technol., vol. 5, no. 1, pp. 1_167, 2012.

[22] X. Wang, K. Liu, and J. Zhao, ``Handling cold-start problem in review spam detection by jointly embedding texts and behaviors,'' in Proc. 55th Annu. Meeting Assoc. Comput. Linguistics, vol. 1, 2017, pp. 366_376.

[23] N. A. Patel and R. Patel, ``A survey on fake review detection using machine learning techniques,'' in Proc. 4th Int. Conf. Comput. Commun. Autom. (ICC), Dec. 2018, pp. 1_6.

[24] C.-C. Wang, M.-Y. Day, C.-C. Chen, and J.-W. Liou, ``Detecting spamming reviews using long short-term memory recurrent neural network framework,'' Proc. 2nd Int. Conf. E-Commerce, E-Business E-Government (ICEEG), 2018, pp. 16_20.

[25] X. Tang, T. Qian, and Z. You, ``Generating behavior features for cold start spam review detection with adversarial learning,'' Inf. Sci., vol. 526, pp. 274_288, Jul. 2020.

© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[26] Z. Li, H. Yao, and F. Ma, ``Learning with small data,'' Proc. 13th Int. Conf. Web Search Data Mining, Jan. 2020, pp. 3539_3540.

[27] L. Li, B. Qin, W. Ren, and T. Liu, ``Document representation and feature combination for deceptive spam review detection,'' Neurocomputing, vol. 254, pp. 33_41, Sep. 2017.

[28] S. Mani, S. Kumari, A. Jain, and P. Kumar, ``Spam review detection using ensemble machine learning,'' in Proc. Int. Conf. Mach. Learn. Data Mining Pattern Recognit. New York, NY, USA: Springer, 2018, pp. 198_209.

[29] Á. Hernández-Castañeda, H. Calvo, A. Gelbukh, and J. J. G. Flores, ``Cross-domain deception detection using support vector networks,'' Soft Comput., vol. 21, no. 3, pp. 585_595, Feb. 2017.

[30] L.-Y. Dong, S.-J. Ji, C.-J. Zhang, Q. Zhang, D. W. Chiu, L.-Q. Qiu, and D. Li, ``An unsupervised topic-sentiment joint probabilistic model for detecting deceptive reviews,'' Expert Syst. Appl., vol. 114, pp. 210_223, Dec. 2018.

[31] U. Aslam, M. Jayabalan, H. Ilyas, and A. Suhail, ``A survey on opinion spam detection methods,'' Int. J. Sci. Technol. Res., vol. 8, no. 9, pp. 1_10, 2019.

[32] W. Liu, W. Jing, andY. Li, ``Incorporating feature representation into BiLSTM for deceptive review detection,'' Computing, vol. 102, pp. 701_715, Nov. 2019.

[33] R. Xu, Y. Xia, K.-F. Wong, and W. Li, ``Opinion annotation in on-line Chinese product reviews,'' in LREC, vol. 8, 2008, pp. 26_30.

[34] A. K Samha, Y. Li, and J. Zhang, ``Aspect-based opinion extraction from customer reviews,'' 2014, arXiv:1404.1982. [Online]. Available: http://arxiv.org/abs/1404.1982

[35] H. Deng, L. Zhao, N. Luo, Y. Liu, G. Guo, X. Wang, Z. Tan, S. Wang, and F. Zhou, ``Semi-supervised learning based fake review detection,'' in Proc. IEEE Int. Symp. Parallel Distrib. Process. Appl. IEEE Int. Conf. Ubiquitous Comput. Commun. (ISPA/IUCC), Dec. 2017, pp. 1278_1280.

[36] X. Tang, T. Qian, and Z. You, ``Generating behavior features for cold start spam review detection,'' in Proc. Int. Conf. Database Syst. Adv. Appl. Chiang Mai, Thailand: Springer, 2019, pp. 324_328.

[37] W. Zhang, Y. Du, T. Yoshida, and Q. Wang, ``DRI-RCNN: An approach to deceptive review identi_cation using recurrent convolutional neural network,'' Inf. Process. Manage., vol. 54, no. 4, pp. 576_592, 2018.

[38] J. Sánchez-Junquera, L. Villaseñor-Pineda, M. Montes-y-Gómez, P. Rosso, and E. Stamatatos, ``Masking domain-speci_c information for cross-domain deception detection,'' Pattern Recognit. Lett., vol. 135, pp. 122_130, Jul. 2020.

[39] C. Visani, N. Jadeja, and M. Modi, ``A study on different machine learning techniques for spam review detection,'' in Proc. Int. Conf. Energy, Commun., Data Anal. Soft Comput. (ICECDS), Aug. 2017, pp. 676_679.

[40] Z. Zhao, C. Li, Y. Zhang, J. Z. Huang, J. Luo, S. Feng, and J. Fan, ``Identifying and analyzing popular phrases multi-dimensionally in social media data,'' Int. J. DataWarehousing Mining, vol. 11, no. 3, pp. 98_112, Jul. 2015.

[41] N. Jain, A. Kumar, S. Singh, C. Singh, and S. Tripathi, ``Deceptive reviews detection using deep learning techniques,'' in Proc. Int. Conf. Appl. Natural Lang. Inf. Syst. Salford, U.K.: Springer, 2019, pp. 79_91.

[42] F. Khurshid, Y. Zhu, C. W. Yohannes, and M. Iqbal, ``Recital of supervised learning on review spam detection: An empirical analysis,'' in Proc. 12th Int. Conf. Intell. Syst. Knowl. Eng. (ISKE), Nov. 2017, pp. 1_6.

[43] C. M. Yilmaz and A. O. Durham, ``SPR2EP: A semi-supervised spam review detection framework,'' in Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM), Aug. 2018, pp. 306_313.

[44] Q. Li, Q.Wu, C. Zhu, J. Zhang, andW. Zhao, ``An inferable representation learning for fraud review detection with the cold-start problem,'' Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2019, pp. 1_8.

[45] S. Noekhah, N. B. Salim, and N. H. Zakaria, ``A novel model for opinion spam detection based on multi-iteration network structure,'' Adv. Sci. Lett., vol. 24, no. 2, pp. 1437_1442, Feb. 2018.

[46] S. Noekhah, N. B. Salim, and N. H. Zakaria, ``Opinion spam detection: Using multi-iterative graph-based model,'' Inf. Process. Manage., vol. 57, no. 1, Jan. 2020, Art. no. 102140.

© 2022 Little Lion Scientific
ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[47] N. N. Ho-Dac, S. J. Carson, and W. L. Moore, ``The effects of positive and negative online customer reviews: Do brand strength and category maturity matter?'' J. Marketing, vol. 77, no. 6, pp. 37_53, Nov. 2013.

[48] F. Zhu and X. Zhang, ``Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics,'' J. Marketing, vol. 74, no. 2, pp. 133_148, Mar. 2010.

[49] J. C. Rodrigues, J. T. Rodrigues, V. L. K. Gonsalves, A. U. Naik, P. Shetgaonkar, and S. Aswell, ``Machine & deep learning techniques for detecting fake reviews: A survey,'' in Proc. Int. Conf. Emerg. Trends Inf. Technol. Eng. (ic-ETITE), Feb. 2020, pp. 1_8.

[50] Z.-Y. Zeng, J.-J. Lin, M.-S. Chen, J.-L. Liu, ``A review structure-based ensemble model for deceptive review spam,'' Information, vol. 10, no. 7, p. 243, Jul. 2019.

[51] R. Mohawesh, S. Tran, R. Ollington, and S. Xu, ``Analysis of concept drift in fake reviews detection,'' Expert Syst. Appl., vol. 169, May 2021, Art. No. 114318.

[52] I. Peñalver-Martinez, F. Garcia-Sanchez, and J. L. Sánchez- Cervantes, ``Feature-based opinion mining through ontologies,'' Expert Syst. Appl., vol. 41, no. 13, pp. 5995_6008, Oct. 2014.

[53] A.-M. Popescu and O. Etzioni, ``Extracting product features and opinions from reviews,'' in Natural Language Processing and Text Mining. London, U.K.: Springer, 2007, pp. 9_28. [54] M. Hu and B. Liu, ``Mining and summarising customer reviews,'' in Proc. ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2004, pp. 168_177.

[54] Y. Tian, M. Mirzabagheri, P. Tirandazi, and S. M. H. Bamakan, ``A nonconvex semi-supervised approach to opinion spam detection by ramp one-class SVM,'' Inf. Process. Manage., vol. 57, no. 6, Nov. 2020, Art. No. 102381.

[55] Q. Li, Q.Wu, C. Zhu, J. Zhang, andW. Zhao, ``Unsupervised user behavior representation for fraud review detection with the cold-start problem,'' Proc. Paci_c-Asia Conf. Knowl. Discovery Data Mining. Springer, 2019, pp. 222_236.

[56] C. Yuan, W. Zhou, Q. Ma, S. Lv, J. Han, and S. Hu, ``Learning review representations from the user and product-level information for spam detection,'' in Proc. IEEE Int. Conf. Data Mining (ICDM), Nov. 2019, pp. 1444_1449.

[57] D. U. Vidanagama, T. P. Silva, and A. S. Karunananda, ``Deceptive consumer review detection: A survey,'' Artif. Intell. Rev., vol. 53, no. 2, pp. 1323_1352, Feb. 2020.

[58] J. Li, P. Lv, W. Xiao, L. Yang, and P. Zhang, ``Exploring groups of opinion spam using sentiment analysis guided by nominated topics,'' Expert Syst. Appl., vol. 171, Jun. 2021, Art. No. 114585.

[59] G. Shan, L. Zhou, and D. Zhang, ``From con_icts and confusion to doubts: Examining review inconsistency for fake review detection,'' Decis. Support Syst., vol. 144, May 2021, Art. No. 113513.

[60] J. Wang, H. Kan, F. Meng, Q. Mu, G. Shi, and X. Xiao, ``Fake review detection based on multiple feature fusion and rolling collaborative training,'' IEEE Access, vol. 8, pp. 182625_182639, 2020.

[61] Y. Ganin and V. Lempitsky, ``Unsupervised domain adaptation by backpropagation,'' in Proc. Int. Conf. Mach. Learn., 2015, pp. 1180_1189.

[62] Y.Wu, E.W. T. Ngai, P.Wu, and C.Wu, ``Fake online reviews: Literature review, synthesis, and directions for future research,'' Decis. Support Syst., vol. 132, May 2020, Art. No. 113280.

[63] L. You, Q. Peng, Z. Xiong, D. He, M. Qiu, and X. Zhang, ``Integrating aspect analysis and local outlier factor for intelligent review spam detection,'' Future Gener. Comput. Syst., vol. 102, pp. 163_172, Jan. 2020.

[64] N. Cao, S. Ji, D. K. W. Chiu, M. He, and X. Sun, ``A deceptive review detection framework: Combination of coarse and _ne-grained features,'' Expert Syst. Appl., vol. 156, Oct. 2020, Art. No. 113465.

[65] J. Yao, Y. Zheng, and H. Jiang, ``An ensemble model for fake online review detection based on data resampling, feature pruning, and parameter optimization,'' IEEE Access, vol. 9, pp. 16914_16927, 2021.

[66] M. Peng, Q. Zhang, Y.-G. Jiang, and X.-J. Huang, ``Cross-domain sentiment classi_cation with target domain speci_c information,'' in Proc. 56th Annu. Meeting Assoc. Comput. Linguistics, vol. 1, 2018, pp. 2505_2513.

[67] A. Ligthart, C. Catal, and B. Tekinerdogan, ``Analyzing the effectiveness of semi-supervised learning approaches for opinion spam classi_cation," Appl. Soft Comput., vol. 101, Mar. 2021, Art. No. 107023.

[68] A. Fahfouh, J. Rif_, M. Adnane Mahraz, H. Tairi, ``PV-DAE: A hybrid model for deceptive opinion spam based on neural network architectures," Expert Syst. Appl., vol. 157, Nov. 2020, Art. No. 113517.

[69] N. Dhamani, P. Azure, J. L. Gleason, and J. Morgan, ``Using deep networks and transfer learning to address disinformation," 2019, arXiv:1905.10412. [Online]. Available: http://arxiv.org/abs/1905.10412

[70] Z. Guo, L. Tang, T. Guo, K. Yu, M. Alazab, and A. Shalaginov, ``Deep graph neural network-based spammer detection under the perspective of heterogeneous cyberspace," Future Gener. Comput. Syst., vol. 117, pp. 205_218, Apr. 2021.

[71] S. Rayana and L. Akoglu, ``Collective opinion spam detection: Bridging review networks and metadata," in Proc. 21th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2015, pp. 985_994.

[72] M. Crawford, T. M. Khoshgoftaar, H. Al Najada, ``Survey of review spam detection using machine learning techniques," J. Big Data, vol. 2, no. 1, p. 23, Dec. 2015.

[73] Y. Ren and D. Ji, ``Learning to detect deceptive opinion spam: A survey," IEEE Access, vol. 7, pp. 42934_42945, 2019.

[74] X. Wang, K. Liu, and J. Zhao, ``Detecting deceptive review spam via attention-based neural networks," in Proc. Nat. CCF Conf. Natural Lang. Process. Chin. Comput. Dalian, China: Springer, 2017, pp. 866_876.

[75] X. Wang, K. Liu, S. He, and J. Zhao, ``Learning to represent review with tensor decomposition for spam detection," in Proc. Conf. Empirical Methods Natural Lang. Process., 2016, pp. 866_875.

[76] M. Ott, Y. Choi, C. Cardie, and J. T. Hancock, ``Finding deceptive opinion spam by any stretch of the imagination," in Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Human Lang. Technol., vol. 1, 2011, pp. 309_319.

[77] S. Kim, H. Chang, S. Lee, M. Yu, and J. Kang, ``Deep semantic frame-based deceptive opinion spam analysis," in Proc. 24th ACM Int. Conf. Inf. Knowl. Manage., Oct. 2015, pp. 1131_1140.

[78] R. Barbados, O. Araque, and C. A. Iglesias, ``A framework for fake review detection in online consumer electronics retailers," Inf. Process. Manage., vol. 56, no. 4, pp. 1234_1244, Jul. 2019.

[79] H. Li, Z. Chen, B. Liu, X. Wei, and J. Shao, ``Spotting fake reviews via collective positive-unlabeled learning," in Proc. IEEE Int. Conf. Data Mining, Dec. 2014, pp. 899_904.

[80] H. Li, G. Fei, S. Wang, B. Liu, J. Shao, ``Bimodal distribution and co-bursting in review spam detection," in Proc. 26th Int. Conf. World Wide Web, Apr. 2017, pp. 1063_1072.

[81] Y. Li, ``Highlighting the fake reviews in review sequence with the suspicious contents and behaviors," J. Inf. Comput.al Sci., vol. 12, no. 4, pp. 1615_1627, Mar. 2015.

[82] A. Mukherjee, A. Kumar, B. Liu R. Ghosh, ``Spotting opinion spammers using behavioral footprints," Proc. 19th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, Aug. 2013, pp. 632_640.

[83] F. H. Li, M. Huang, Y. Yang, and X. Zhu, ``Learning to identify review spam," in Proc. 22nd Int. Joint Conf. Artif. Intell., 2011, pp. 1_6.

[84] N. Jindal and B. Liu, ``Review spam detection," in Proc. 16th Int. Conf. World Wide Web, 2007, pp. 1189_1190.

[85] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, ``Exploiting burstiness in reviews for review spammer detection," in Proc. 7th Int. AAAI Conf. Weblogs Social Media, 2013, pp. 1_10.

[86] M. Jiang, P. Cui, and C. Faloutsos, ``Suspicious behavior detection: Current trends and future directions," IEEE Intell. Syst., vol. 31, no. 1, pp. 31_39, Jan. 2016.

[87] S. P. Algur, N.Ayachit, and J. G. Biradar, ``Exponential distribution model for review spam detection," Int. J. Adv. Res. Comput. Sci., vol. 8, no. 3, 2017.

[88] T. Fornaciari and M. Poesio, ``Identifying fake Amazon reviews as learning from crowds," in Proc. 14th Conf. Eur. Chapter Assoc. Comput. Linguistics, 2014, pp. 279_287.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[89] A. S. Abu Hammad, ``An approach for detecting spam in Arabic opinion reviews,'' in An Approach for Detecting Spam in Arabic Opinion Reviews. Amman, Jordan: The International Arab Journal of Information Technology, 2013.

[90] A. Mukherjee, B. Liu, and N. Glance, ``Spotting fake reviewer groups in consumer reviews,'' in Proc. 21st Int. Conf. World Wide Web (WWW), 2012, pp. 191_200.

[91] E.-P. Lim, V.-A. Nguyen, N. Jindal, B. Liu, and H. W. Lauw, ``Detecting product review spammers using rating behaviors,'' Proc. 19th ACM Int. Conf. Inf. Knowl. Manage. (CIKM), 2010, pp. 939_948.

[92] Y. Ren, D. Ji, and H. Zhang, ``Positive unlabeled learning for deceptive reviews detection,'' in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 488_498.

[93] S.-J. Ji, Q. Zhang, J. Li, S. Xu, L. Yi, and M. Gong, ``A burst-based unsupervised method for detecting review spammer groups,'' Inf. Sci., vol. 536, pp. 454_469, Oct. 2020.

[94] A. Rastogi, M. Mehrotra, and S. S. Ali, ``Effective opinion spam detection: A study on review metadata versus content,'' J. Data Inf. Sci., vol. 5, no. 2, pp. 76_110, Apr. 2020.

[95] J. Serrano-Guerrero, J. A. Olivas, E. Herrera-Viedma, ``Sentiment analysis: A review and comparative analysis of Web services,'' Inf. Sci., vol. 311, pp. 18_38, Aug. 2015.

[96] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, ``What yelp fake review_lter might be doing?'' in Proc. 7th Int. AAAI Conf.Weblogs Social Media, 2013, pp. 1_10.

[97] A. U. Akram, H. U. Khan, M. Sha_, ``Finding rotten eggs: A review spam detection model using diverse feature sets,'' KSII Trans. Internet Inf. Syst., vol. 12, no. 10, pp. 5120_5142, 2018.

[98] M. Ott, C. Cardie, and J. T. Hancock, ``Negative deceptive opinion spam,'' in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol., 2013, pp. 497_501.

[99] J. Li, M. Ott, C. Cardie, and E. Hovy, ``Towards a general rule for identifying deceptive opinion spam,'' in Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics, vol. 1, 2014, pp. 1566_1576.

[100] T. Ong, M. Mannino, and D. Gregg, ``Linguistic characteristics of shill reviews,'' Electron. Commerce Res. Appl., vol. 13, no. 2, pp. 69_78, Mar. 2014.

[101] P. Rayson, A. Wilson, and G. Leech, ``Grammatical word class variation within the British National Corpus sampler,'' in New Frontiers Corpus Research. Amsterdam, The Netherlands: Brill | Rodopi, 2002, pp. 295_306.

[102] D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan, Longman Grammar of Spoken and Written English. London, U.K.: Longman, 2000.

[103] F. Mairesse, M. A. Walker, M. R. Mehl, and R. K. Moore, ``Using linguistic cues for the automatic recognition of personality in conversation and text,'' J. Artif. Intell. Res., vol. 30, pp. 457_500, Nov. 2007.

[104] W. L. Cade, B. A. Lehman, and A. Olney, ``An exploration of off-topic conversation,'' in Proc. Annu. Conf. North Amer. Chapter Assoc. Comput. Linguistics, 2010, pp. 669_672.

[105] A. Vrij, S. Mann, S. Kristen, and R. P. Fisher, ``Cues to deception and ability to detect lies as a function of police interview styles,'' Law Hum. Behav., vol. 31, no. 5, pp. 499_518, 2007.

[106] S. Shojaee, M. A. A. Murad, S. Nadali, ``Detecting deceptive reviews using lexical and syntactic features,'' in Proc. 13th Int. Conf. Intelligent Syst. Design Appl., Dec. 2013, pp. 53_58.

[107] Y. Bengio, R. Ducharme, P. Vincent, and C. Janvin, ``A neural probabilistic language model,'' J. Mach. Learn. Res., vol. 3, pp. 1137_1155, Feb. 2003.

[108] W. Y. Zou, R. Socher, D. Cer, and C. D. Manning, ``Bilingual word embeddings for phrase-based machine translation,'' in Proc. Conf. Empir- ical Methods Natural Lang. Process., 2013, pp. 1393_1398.

[109] A. Bordes, S. Chopra, and J. Weston, ``Question answering with subgraph embeddings,'' 2014, arXiv:1406.3676. [Online]. Available: http://arxiv.org/abs/1406.3676

[110] D. Weiss, C. Alberti, M. Collins, and S. Petrov, ``Structured training for neural network transition-based parsing,'' 2015, arXiv:1506.06158. [Online]. Available: http://arxiv.org/abs/1506.06158

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[111] G. Salton, A.Wong, and C. S. Yang, ``A vector space model for automatic indexing,'' Commun. ACM, vol. 18, no. 11, pp. 613_620, Nov. 1975.

[112] R. Collobert and J. Weston, ``A uni_ed architecture for natural language processing: Deep neural networks with multitask learning,'' in Proc. 25th Int. Conf. Mach. Learn., 2008, pp. 160_167.

[113] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, ``Distributed representations of words and phrases and their compositionality,'' in Proc. Adv. Neural Inf. Process. Syst., 2013, pp. 3111_3119.

[114] W. Ling, T. Luís, L. Marujo, I. Trancoso, ``Finding function in form: Compositional character models for open vocabulary word representation,'' 2015, arXiv:1508.02096. [Online]. Available: http://arxiv.org/abs/1508.02096

[115] J. Pennington, R. Socher, and C. Manning, ``Glove: Global vectors for word representation,'' in Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP), 2014, pp. 1532_1543.

[116] M. Baroni, G. Dinu, and G. Kruszewski, ``Don't count, predict! A systematic comparison of context-counting vs. context-predicting semantic vectors,'' in Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics, vol. 1, 2014, pp. 238_247.

[117] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, ``Bag of tricks for ef_cient text classi_cation,'' 2016, arXiv:1607.01759. [Online]. Available: http://arxiv.org/abs/1607.01759 [119]

[118] Q. Le and T. Mikolov, ``Distributed representations of sentences and documents,'' in Proc. Int. Conf. Mach. Learn., 2014, pp. 1188_1196.

[119] S. Achsah and E. H. Nfaoui, ``Language representation learning models: A comparative study,'' in Proc. 13th Int. Conf. Intell. Syst.: Theories Appl., 2020, pp. 1_7.

[120] A. Grover and J. Leskovec, ``node2vec: Scalable feature learning for networks,'' in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 855_864.

[121] F. Almeida and G. Xexéo, ``Word embeddings: A survey,'' 2019, arXiv:1901.09069. [Online]. Available: http://arxiv.org/abs/1901.09069

[122] S. Feng, R. Banerjee, and Y. Choi, ``Syntactic stylometry for deception detection,'' in Proc. 50th Annu. Meeting Assoc. Comput. Linguistics, vol. 2, 2012, pp. 171_175.

[123] Y. Ren, L. Yin, and D. Ji, ``Deceptive reviews detection based on language structure and sentiment polarity,'' J. Frontiers Comput. Sci. Technol., vol. 8, no. 3, pp. 313_320, 2014.

[124] D. Tao, Y. Guo, Y. Li, and X. Gao, ``Tensor rank preserving discriminant analysis for facial recognition,'' IEEE Trans. Image Process., vol. 27, no. 1, pp. 325_334, Jan. 2018.

[125] X. Zhang, X. Liu, and Z. J. Wang, ``Evaluation of a set of new ORF kernel functions of SVM for speech recognition,'' Eng. Appl. Artif. Intell., vol. 26, no. 10, pp. 2574_2580, Nov. 2013.

[126] D. Tao, X. Lin, L. Jin, and X. Li, ``Principal component 2-D long short-term memory for font recognition on single Chinese characters,'' IEEE Trans. Cybern., vol. 46, no. 3, pp. 756_765, Mar. 2016.

[127] J. Z. Lei and A. A. Ghorbani, ``Improved competitive learning neural networks for network intrusion and fraud detection,'' Neurocomputing, vol. 75, no. 1, pp. 135_145, Jan. 2012.

[128] L. Nanni, ``An ensemble of classi_ers for the diagnosis of erythematosquamous diseases,'' Neurocomputing, vol. 69, nos. 7_9, pp. 842_845, Mar. 2006.

[129] M. Al-Hawawreh and E. Sitnikova, ``Leveraging deep learning models for ransomware detection in the Industrial Internet of Things environment,'' in Proc. Mil. Commun. Inf. Syst. Conf. (MilCIS), Nov. 2019, pp. 1_6.

[130] M. Al-Hawawreh and E. Sitnikova, ``Industrial Internet of Things based ransomware detection using stacked variational neural network,'' in Proc. 3rd Int. Conf. Big Data Internet Things (BDIOT), 2019, pp. 126_130.

[131] P. P. K. Chan, C. Yang, D. S. Yeung, andW.W. Y. Ng, ``Spam _ltering for short messages in an adversarial environment,'' Neurocomputing, vol. 155, pp. 167_176, May 2015.

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

[132] R. M. Silva, T. C. Alberto, T. A. Almeida, and A. Yamakami, ``Towards _ltering undesired short text messages using an online learning approach with semantic indexing,'' Expert Syst. Appl., vol. 83, pp. 314_325, Oct. 2017.

[133] C. Hua Li and J. Xiangji Huang, ``Spam _ltering using semantic similarity approach and adaptive BPNN,'' Neurocomputing, vol. 92, pp. 88_97, Sep. 2012.

[134] T. C. Alberto, J. V. Lochter, and T. A. Almeida, ``Post or block? Advances in automatically _ltering undesired comments,'' J. Intell. Robotic Syst., vol. 80, no. S1, pp. 245_259, Dec. 2015.

[135] J. Eisenstein, A. Ahmed, and E. P. Xing, ``Sparse additive generative models of text,'' in Proc. 28th Int. Conf. Mach. Learn., 2011, pp. 1041_1048.

[136] B. M. DePaulo, J. J. Lindsay, B. E. Malone, L. Muhlenbruck, K. Charlton, and H. Cooper, ``Cues to deception,'' Psychol. Bull., vol. 129, no. 1, p. 74, 2003.

[137] V. Pérez-Rosas and R. Mihalcea, ``Cross-cultural deception detection,'' in Proc. 52nd Annu. Meeting Assoc. Comput. Linguistics, vol. 2, 2014, pp. 440_445.

[138] J. Sánchez-Junquera, L. Villaseñor-Pineda, and P. Rosso, ``Character N-grams for detecting deceptive controversial opinions,'' in Proc. Int. Conf. Cross-Lang. Eval. Forum Eur. Lang. Avignon, France: Springer, 2018, pp. 135_140.

[139] L. Cagnina and P. Rosso, ``Classi_cation of deceptive opinions using a low dimensionality representation,'' in Proc. 6th Workshop Comput. Approaches Subjectivity, Sentiment Social Media Anal., 2015, pp. 58_66.

[140] R. Mihalcea and C. Strapparava, ``The lie detector: Explorations in the automatic recognition of deceptive language,'' in Proc. ACL-IJCNLP Conf. Short Papers, 2009, pp. 309_312.

[141] S. Nilizadeh, H. Aghakhani, E. Gustafson, and G. Vigna, ``Think outside the dataset: Finding fraudulent reviews using cross dataset analysis,'' Proc. World Wide Web Conf., 2019, pp. 3108_3115.

[142] J. Li, C. Cardie, and S. Li, ``Topicspam: Atopic-model based approach for spam detection,'' in Proc. 51st Annu. Meeting Assoc. Comput. Linguistics, vol. 2, 2013, pp. 217_221.

[143] E. A. Suess and B. E. Trumbo, ``Using Gibbs samplers to compute Bayesian posterior distributions,'' in Introduction to Probability Simulation and Gibbs Sampling with R. New York, NY, USA: Springer, 2010, pp. 219_248.

[144] C. Elkan and K. Noto, ``Learning classi_ers from only positive and unlabeled data,'' in Proc. 14th ACMSIGKDD Int. Conf. Knowl. Discovery Data Mining (KDD), 2008, pp. 213_220.

[145] H. Zhao, Z. Lu, and P. Poupart, ``Self-adaptive hierarchical sentence model,'' 2015, arXiv:1504.05070. [Online]. Available: http://arxiv.org/abs/1504.05070

[146] B. Liu, X. Li, ``Partially supervised classi_cation of text documents,'' in Proc. ICML, vol. 2, 2002, pp. 387_394.

[147] T. Joachims, ``Transductive inference for text classi_cation using support vector machines,'' in Proc. ICML, vol. 99, 1999, pp. 200_209.

[148] Y. Zhang and D.-Y. Yeung, ``A convex formulation for learning task relationships in multi-task learning,'' presented at the 26th Conf. Uncertainty Artif. Intell., Catalina Island, CA, USA, 2010.

[149] G. Wang, S. Xie, B. Liu, and P. S. Yu, ``Review graph-based online store review spammer detection,'' in Proc. IEEE 11th Int. Conf. Data Mining, Dec. 2011, pp. 1242_1247.

[150] F. Gieseke, A. Airola, T. Pahikkala, and O. Kramer, ``Sparse quasi-Newton optimization for semi-supervised support vector machines,'' in Proc. ICPRAM, vol. 1, 2012, pp. 45_54.

[151] R. Collobert, J. Weston, and P. Kuksa, ``Natural language processing (almost) from scratch,'' J. Mach. Learn. Res., vol. 12 pp. 2493_2537, Aug. 2011.

[152] Y. Jing, ``Research of deceptive opinion spam recognition based on deep learning,'' East China Normal Univ., Shanghai, China, Tech. Rep., 2014.

[153] Y. Ren and D. Ji, ``Neural networks for deceptive opinion spam detection: An empirical study,'' Inf. Sci., vols. 385_386, pp. 213_224, Apr. 2017.

[154] Y. Ren, R.Wang, and D. Ji, ``A topic-enhanced word embedding for Twitter sentiment classi_cation,'' Inf. Sci., vol. 369, pp. 188_198, Nov. 2016.

[155] S. M. Asadullah and S. Viraktamath, ``Classi_cation of Twitter spam based on pro_le and message model using SVM,'' Dept. Electron. Commun., SDM College Eng. Technol., Karnataka, India, Tech. Rep., 2017.

[156] S. Seneviratne, A. Seneviratne, M. A. Kaafar, ``Spam mobile apps: Characteristics, detection, and in the wild analysis,'' ACM Trans. Web, vol. 11, no. 1, p. 4, 2017.

[157] H.-H. Huang, Y.-W. Wen, and H.-H. Chen, ``Detection of false online advertisements with DCNN,'' in Proc. 26th Int. Conf. World Wide Web Companion (WWW Companion), 2017, pp. 795_796.

[158] A. Bordes, N. Usunier, A. Garcia-Duran,``Translating embeddings for modeling multi-relational data,'' in Proc. Adv. Neural Inf. Process. Syst., 2013, pp. 2787_2795.

[159] S. Lai, L. Xu, K. Liu, and J. Zhao, ``Recurrent convolutional neural networks for text classi_cation,'' in Proc. 29th AAAI Conf. Artif. Intell., 2015, pp. 1_7.

[160] V. W. Feng and G. Hirst, ``Detecting deceptive opinions with pro_le compatibility,'' in Proc. 6th Int. Joint Conf. Natural Lang. Process., 2013, pp. 338_346.

[161] Y. Ren, Y. Zhang, M. Zhang, and D. Ji, ``Context-sensitive Twitter sentiment classi_cation using neural network,'' in Proc. 13th AAAI Conf. Artif. Intell., 2016, pp. 215_221.

[162] J. Li, M. Ott, and C. Cardie, ``Identifying manipulated offerings on review portals,'' in Proc. Conf. Empirical Methods Natural Lang. Process., 2013, pp. 1933_1942.

[163] F. Gräÿer, S. Kallumadi, H. Malberg, and S. Zaunseder, ``Aspect-based sentiment analysis of drug reviews applying cross-domain and cross-data learning,'' in Proc. Int. Conf. Digit. Health, Apr. 2018, pp. 121_125.

[164] A. L. Maas, R. E. Daly, ``Learning word vectors for sentiment analysis,'' in Proc. 49th Annu. Meeting Assoc. Comput. Linguistics: Hum. Lang. Technol., vol. 1, 2011, pp. 142_150.

[165] Z. Pan, W. Yu, X. Yi, A. Khan, F. Yuan, and Y. Zheng, ``Recent progress on generative adversarial networks (GANs): A survey,'' IEEE Access, vol. 7, pp. 36322_36333, 2019.

[166] I. Goodfellow, ``Generative adversarial nets,'' in Proc. Adv. Neural Inf. Process. Syst., 2014, pp. 2672_2680.

[167] M. Nickel, V. Tresp, and H.-P. Kriegel, ``A three-way model for collective learning on multi-relational data,'' in Proc. ICML, vol. 11, 2011, pp. 809_816.

[168] J. Li, ``Feature weight tuning for recursive neural networks,'' 2014, arXiv:1412.3714. [Online]. Available: http://arxiv.org/abs/1412.3714

[169] Z. Wang, J. Zhang, J. Feng, and Z. Chen, ``Knowledge graph embedding by translating on hyperplanes,'' in Proc. 28th AAAI Conf. Artif. Intell., 2014, pp. 1112_1119.

[170] Y.-R. Chen and H.-H. Chen, ``Opinion spam detection in Web forum: A real case study,'' in Proc. 24th Int. Conf. World Wide Web, 2015, pp. 173_183.

[171] S. K. Maity, K. C. Santosh, and A. Mukherjee, ``Spam2vec: Learning biased embeddings for spam detection in Twitter,'' in Proc. Companion Web Conf., 2018, pp. 63_64.

[172] F. Xia, J. Liu, H. Nie, Y. Fu, ``Random walks: A review of algorithms and applications,'' IEEE Trans. Emerg. Topics Comput. Intell., vol. 4, no. 2, pp. 95_107, Apr. 2020.

[173] X. Wang, W. Jiang, and Z. Luo, ``Combination of convolutional and recurrent neural network for sentiment analysis of short texts,'' in Proc. 26th Int. Conf. Comput. Linguistics: Tech. Papers (COLING), 2016, pp. 2428_2437.

[174] X. Zhang, J. Zhao, and Y. LeCun, ``Character-level convolutional networks for text classi_cation,'' 2015, arXiv:1509.01626. [Online]. Available: http://arxiv.org/abs/1509.01626 [176] Z. Yang, D. Yang, C. Dyer, X. He, A. Smola, and E. Hovy, ``Hierarchical attention networks for document classi_cation,'' in Proc. Conf. North Amer. Chapter Assoc. Comput. Linguistics: Hum. Lang. Technol., 2016, pp. 1480_1489.

[175] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, ``BERT: Pretraining of deep

bidirectional transformers for language understanding," 2018, arXiv:1810.04805. [Online]. Available: http://arxiv.org/abs/1810. 04805

[176] [178] V. Sanh, L. Debut, J. Chaumont, and T. Wolf, ``DistilBERT, a distilled version of BERT: Smaller, faster, cheaper and lighter," 2019, arXiv:1910.01108. [Online]. Available: http://arxiv.org/abs/1910. 01108

[177] Y. Liu, M. Ott, N. Goyal. Zettlemoyer, and V. Stoyanov, ``Roberta: A robustly optimized BERT pretraining approach," 2019, arXiv:1907.11692. [Online]. Available: http://arxiv.org/abs/1907.11692

[178] G. Vallone and L. Longo, ``Explainable arti_cial intelligence: A systematic review," 2020, arXiv:2006.00093. [Online]. Available: http://arxiv.org/abs/2006.00093

[179] J. Zhao, N. Cao, Z.Wen, Y. Song, Y.-R. Lin, and C. Collins, ``#FluxFlow: Visual analysis of anomalous information spreading on social media," IEEE Trans. Vis. Comput. Graph., vol. 20, no. 12, pp. 1773_1782, Dec. 2014.

[180] S. S. Khan and M. G. Madden, ``A survey of recent trends in one class classi_cation," Proc. Irish Conf. Artif. Intell. Cogn. Sci. Dublin, Ireland: Springer, 2009, pp. 188_197.

[181] R. Chalapathy, A. Krishna Menon, and S. Chawla, ``Anomaly detection using one-class neural networks," 2018, arXiv:1802.06360. [Online]. Available: http://arxiv.org/abs/1802.06360

[182] S. S. Khan and M. G. Madden, ``One-class classi_cation: Taxonomy of study and review of techniques," Knowl. Eng. Rev., vol. 29, no. 3, pp. 345_374, Jun. 2014.

[183] C. Désir, S. Bernard, C. Petitjean, and L. Heutte, ``One class random forests," Pattern Recognit., vol. 46, no. 12, pp. 3490_3506, Dec. 2013.

[184] Y. Ganin and V. Lempitsky, ``Unsupervised domain adaptation by backpropagation," 2014, arXiv:1409.7495. [Online]. Available: http://arxiv.org/abs/1409.7495

[185] P. Capuozzo, I. Lauriola, ``DecOp: A multilingual and multi-domain corpus for detecting deception in typed text," in Proc. 12th Lang. Resour. Eval. Conf., 2020, pp. 1423_1430.

[186] Jindal, N., & Liu, B. (2007a). Analyzing and detecting review spam. Proceedings - IEEE International Conference on Data Mining, ICDM, 547–552. DOI: 10.1109/ICDM.2007 .68

[187] Jindal, N., & Liu, B. (2008). Opinion spam and analysis. Proceedings of the international conference on web search and web data mining 2008, 219–230. DOI: 10.1145/1341531 .1341560

[188] Lai, C. L., Xu, K. Q., Lau, R. Y. K., Li, Y., & Jing, L. (2010). Toward a language modeling approach for consumer review spam detection. Proceedings - IEEE International Conference on E-Business Engineering, ICEBE 2010, 1–8. DOI: 10.1109/ICEBE.2010.47

[189] Algur, S., Hiremath, E., Patil, A., & Shivashankar, S. (2010). Spam detection of customer reviews from web pages. In Proceedings of the 2nd international conference on it and business intelligence (pp. 1–13)

[190] Li, F., Huang, M., Yang, Y., & Zhu, X. (2011). Learning to identify review spam. Ijcai proceedings-international joint conference on artificial intelligence (Vol. 22, p. 2488).

[191] Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013a). Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews. UIC-CS-03-2013. Technical Report.

[192] Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. (2013b). What Yelp Might Fake Review Filter Be Doing? Seventh International AAAI . . . , 409–418.

[193] Li, H., Liu, B., Mukherjee, A., & Shao, J. (2014). Spotting fake reviews using positive-unlabeled learning. Computación y Sistemas, 18(3), 467–475. DOI: https://dx.doi.org/10.13053/ CDS-18-3-2035

[194] Lin, Y., Zhu, T., Wang, X., Zhang, J., & Zhou, A. (2014). Towards Online Review Spam Detection. Proceedings of the 23rd International Conference on World Wide Web, 341– 342.

[195] Istiaq Ahsan, M., Nahian, T., All Kafi, A., Ismail Hossain, M., & Muhammad Shah, F. (2016). An Ensemble Approach to detect Review Spam using hybrid Machine Learning Technique. Computer and Information Technology (ICCIT) 19th International

Conference on IEEE, 388–394. DOI: 10.1109/ICCITECHN.2016.7860229

[196] Zhang, D., Zhou, L., Kehoe, J. L., & Kilic, I. Y. (2016). What Online Reviewer Behaviors Matter? Effects of Verbal and Nonverbal Behaviors on Detection of Fake Online Reviews. Journal of Management Information Systems, 33(2), 456–481. DOI: 10.1080/ 07421222.2016.1205907

[197] M. Kang, J. Ahn, and K. Lee, "Opinion mining using ensemble text hidden Markov models for text classification," Expert Systems with Applications, vol. 94, pp. 218–227, Mar. 2018.

[198] Y. Li et al., "Incorporating knowledge into a neural network for text representation," Expert Systems with Applications, vol. 96, pp. 103–114, Apr. 2018.

[199] E. Teso, M. Olmedilla, "Application of text mining techniques to the analysis of discourse in eWOM communications from a gender perspective," Technological Forecasting and Social Change, vol. 129, pp. 131– 142, Apr. 2018.

[200] E. S. Tellez, S. Miranda-Jiménez, "A case study of Spanish text transformations for Twitter sentiment analysis," Expert Systems with Applications, vol. 81, pp. 457–471, Sep. 2017.

[201] M. Labani, P. Moradi, F. Ahmadizar, and M. Jalili, "A novel multivariate filter method for feature selection in text classification problems," Engineering Applications of Artificial Intelligence, vol. 70, pp. 25–37, Apr. 2018.

[202] J. L. Ong Hui, G. K. Hoon, and W. M. N. Wan Zainon, "Effects of Word Class and Text Position in Sentiment-based News Classification," Procedia Computer Science, vol. 124, pp. 77–85, Jan. 2017.

[203] B. Desmet and V. Hoste, "Online suicide prevention through optimized text classification," Information Sciences, vol. 439–440, pp. 61–78, May 2018.

[204] Y. Du, J. Liu, W. Ke, and X. Gong, "Hierarchy construction and text classification based on the relaxation strategy and least information model," Expert Systems with Applications, vol. 100, pp. 157–164, Jun. 2018.

[205] W. Hadi, Q. A. Al-Radaideh, and S. Alhawari, "Integrating associative rule-based classification with Naïve Bayes for text classification," Applied Soft Computing, vol. 69, pp. 344–356, Aug. 2018.

[206] A. Onan, S. Korukoğlu, and H. Bulut, "Ensemble of keyword extraction methods and classifiers in text classification," Expert Systems with Applications, vol. 57, pp. 232–247, Sep. 2016.

[207] T. Sabbah et al., "Modified frequency-based term weighting schemes for text classification," Applied Soft Computing, vol. 58, pp. 193–206, Sep. 2017.

[208] M. Pavlinek and V. Podgorelec, "Text classification method based on self-training and LDA topic models," Expert Systems with Applications, vol. 80, pp. 83–93, Sep. 2017.

[209] A. Rehman, K. Javed, and H. A. Babri, "Feature selection based on a normalized difference measure for text classification," Information Processing & Management, vol. 53, no. 2, pp. 473–489, Mar. 2017.

[210] R. Wongso, F. A. Lucinda, "News Article Text Classification in Indonesian Language," Procedia Computer Science, vol. 116, pp. 137–143, Jan. 2017.

[211] Y. M. Goh and C. U. Ubeynarayana, "Construction accident narrative classification: An evaluation of text mining techniques," Accident Analysis & Prevention, vol. 108, pp. 122–130, Nov. 2017

[212] Sowmya, B.; Srinivasa, K. Large scale multi-label text classification of a hierarchical data set using Rocchio algorithm. In Proceedings of the 2016 International Conference on Computation System and Information Technology for Sustainable Solutions (CSITSS), Bangalore, India, 6–8 October 2016; pp. 291–296.

[213] Bloehdorn, S.; Hotho, A. Boosting for text classification with semantic features. In International Workshop on Knowledge Discovery on the Web; Springer: Berlin/Heidelberg, Germany, 2004; pp. 149–166.

[214] Genkin, A.; Lewis, D.D.; Madigan, D. Large-scale Bayesian logistic regression for text categorization. Technometrics 2007, 49, 291–304. [CrossRef]

[215] Kim, S.B.; Han, K.S.; Rim, H.C.; Myaeng, S.H. Some effective techniques for naive Bayes text classification.IEEE Trans. Knowl. Data Eng. 2006, 18, 1457–1466.

[216] Chen, K.; Zhang, Z.; Long, J.; Zhang, H. Turning from TF-IDF to TF-IGM for term weighting in text classification. Expert Syst. Appl. 2016, 66, 245–260.

[217] Lodhi, H.; Saunders, C.; Shawe-Taylor, J,Text classification using string kernels. J. Mach. Learn. Res. 2002, 2, 419–444.

[218] Chen, T.; Xu, R.; He, Y.; Wang, X. Improving sentiment analysis via sentence type classification using BiLSTM-CRF and CNN. Expert Syst. Appl. 2017, 72, 221–230.

[219] Yang, Z.; Yang, D.; Dyer, C.; Hierarchical Attention Networks for Document Classification. In Proceedings of the HLT-NAACL, San Diego, CA, USA, 12–17 June 2016; pp. 1480–1489.

[220] Chen, J.; Yan, S.; Wong, K.C. Verbal aggression detection on Twitter comments: Convolutional neural network for short-text sentiment analysis. Neural Comput. Appl. 2018, 1–10.

[221] Jiang, M.; Liang, Y.; Feng, X.; Guan, R. Text classification based on deep belief network and softmax regression. Neural Comput. Appl. 2018, 29, 61–70.

[222] Zhang, X.; Zhao, J.; LeCun, Y. Character-level convolutional networks for text classification. Adv. Neural Inf. Process. Syst. 2015, 28, 649–657.

[223] Kowsari, K.; Heidarysafa, L.E. RMDL: Random Multimodel Deep Learning for Classification. In Proceedings of the 2018 International Conference on Information System and Data Mining, Lakeland, FL, USA, 9–11 April 2018; doi:10.1145/3206098.3206111.

[224] Kowsari, K.; Brown, D.E.; Barnes, L.E. HDLTex: Hierarchical Deep Learning for Text Classification. Machine Learning and Applications (ICMLA). In Proceedings of the 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), Cancun, Mexico, 18–21 December 2017.

[225] Dhamani N, Azunre P, Gleason JL, Corcoran C, Honke G, Kramer S, Morgan J (2019) Using deep networks and transfer learning to address disinformation. arXiv preprint arXiv:1905.10412

[226] Popat K, Mukherjee S, Yates A, Weikum G (2018) Declare: debunking fake news and false claims using evidence-aware deep learning. arXiv preprint arXiv:1809.06416

[227] Wu L, Rao Y, Yu H, Wang Y, Nazir A (2018) False information detection on social media via a deep hybrid model. In: International conference on social informatics, Springer, pp 323–333

[228] Zhang Q, Lipani A, Liang S, Yilmaz E (2019) Reply-aided detection of misinformation via Bayesian deep learning. In: The worldwide web conference, pp 2333–2343

[229] Shu K, Cui L, Wang S, Lee D, Liu H (2019a) defend Explainable fake news detection. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery and data mining, pp 395–405

[230] Chen T, Li X, Yin H, Zhang J (2018) Call attention to rumors: Deep attention-based recurrent neural networks for early rumor detection. In: Pacific-Asia conference on knowledge discovery and data mining, Springer, pp 40–52

[231] Liu Q, Yu F, Wu S, Wang L (2018) Mining significant microblogs for misinformation identification: an attention-based approach. ACM Trans Intell Syst Technol 9:1–20

[232] Popat K, Mukherjee S, Yates A, Weikum G (2018) Declare: debunking fake news and false claims using evidence-aware deep learning. arXiv preprint arXiv:1809.06416

[233] Ruchansky N, Seo S, Liu Y (2017) Csi: A deep hybrid model for fake news detection. In: Proceedings of the 2017 ACM on a conference on information and knowledge management, ACM, pp 797–806

[234] Jia Y, Song X, Zhou J, Liu L, Nie L, Rosenblum DS (2016) Fusing social networks with deep learning for volunteerism tendency prediction. In: Thirtieth AAAI conference on artificial intelligence

[235] Yu F, Liu Q, Wu S, Wang L, Tan T et al. (2017a) A convolutional approach for misinformation identification

[236] Guo H, Cao J, Zhang Y, Guo J, Li J (2018) Rumor detection with hierarchical social

ISSN: 1992-8645 www.jatit.org E-ISSN: 1817-3195

attention network. In: Proceedings of the 27th ACM international conference on information and knowledge management, ACM, pp 943–951

[237] Jin Z, Cao J, Guo H, Zhang Y, Luo J (2017b) Multimodal fusion with recurrent neural networks for rumor detection on microblogs. In: Proceedings of the 25th ACM international conference on multimedia, pp 795–816

[238] Li L, Cai G, Chen N (2018a) A rumor events detection method based on deep bidirectional GRU neural network. In: 2018 IEEE 3rd international conference on the image. Vision and computing (ICIVC), IEEE, pp 755–759

[239] Liu Y, Wu YFB (2018) Early detection of fake news on social media through propagation path classification with recurrent and convolutional networks. In: Thirty-second AAAI conference on artificial intelligence

[240] Ma J, Gao W, Mitra P, Kwon S, Jansen BJ, Wong KF, Cha M (2016) Detecting rumors from microblogs with recurrent neural networks. In: Ijcai, pp 3818–3824

[241] Nguyen TN, Li C, Niederée C (2017b) On early-stage debunking rumors on Twitter: Leveraging the wisdom of weak learners. In: International conference on social informatics, Springer, pp 141–158