IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received January 17, 2022, accepted February 10, 2022, date of publication February 18, 2022, date of current version March 10, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3152806

The Effect of Fake Reviews on e-Commerce During and After Covid-19 Pandemic: SKL-Based Fake Reviews Detection

HINA TUFAIL[1], M. USMAN ASHRAF[2], KHALID ALSUBHI[3], AND HANI MOAITEQ ALJAHDALI[4]

[1]Department of Computer Science, University of Management and Technology, Sialkot 51040, Pakistan
[2]Department of Computer Science, GC Women University Sialkot, Sialkot 51310, Pakistan
[3]Department of Computer Science, King Abdulaziz University, Jeddah 21589, Saudi Arabia
[4]Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah 21589, Saudi Arabia

Corresponding authors: Hina Tufail (hina.tufail@skt.umt.edu.pk) and M. Usman Ashraf (usman.ashraf@gcwus.edu.pk)

This work was supported by the Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, under Grant D-213-611-1443. The authors, therefore, gratefully acknowledge the DSR technical and financial support.

ABSTRACT The outbreak of Covid-19 and the enforcement of lockdown, social distancing, and other precautionary measures lead to a global increase in online shopping. The increasing significance of online shopping and extensive use of e-commerce has increased competition between companies for online selling. Highlights that online reviews play a significant role in boosting a business or slandering it. Product review is an essential factor in customers' decision-making, leading to an intense topic known as fraudulent or fake reviews detection. Given these reviews' power over a business, the treacherous acts of giving false reviews for personal gains have increased with time. In our research, we proposed a fake review detection model by using Text Classification and techniques related to Machine Learning. We used classifiers such as Support Vector Machine, K-Nearest Neighbor, and logistic regression (SKL), using a bigram model that detects fraudulent reviews based on the number of pronouns, verbs, and sentiments. Our proposed methodology for detecting fake online reviews outperforms on the yelp dataset and the TripAdvisor dataset compared to other state-of-the-art techniques with 95% and 89.03% accuracy.

INDEX TERMS Fake reviews, K-Nearest Neighbor (KNN), machine learning, natural language processing, sentiment analysis, support vector machine (SVM).

I. INTRODUCTION

The global pandemic of Covid-19 at the start of the year 2020 leaves a significant impact on everything and everyone. This outbreak shakes the world and shifts the dynamics of e-commerce and online shopping. The enforcement of lockdown and social distancing lead the world to buy products online. One of the most pressing issues faced today is fraud regarding customers' opinions on online products or services relevant to a brand or an organization [1], [2]. The matter has become more sophisticated and organized due to the profit achieved by such pursuit. This phenomenon is called ''Opinion Spamming'' [3], [4]. Dissimilar to other spam, opinion spam are a tad hard to detect as understanding the context is important to detect the deceptiveness of a review. These reviews are posted by people who are inexperienced with the subject, which is why they are considered spam. Given the dynamic nature of the reviews, supervised learning techniques suffer from a few limitations. [2], [5], [6]. Not until the ''quality'' of the review is known, a garbage-in-garbage-out [7] situation can transpire. In a study, [7] it was accentuated by the researchers that fake or genuine reviews are hard to label by humans. This complicates the search for the ground truth for given instances accurately. Due to the versatile nature of these reviews and the lack of reliable data, according to the study [8], [9] methods were utilized to detect deceptive spam. Semi-Supervised techniques were used to improve classification [1], [7], [10]–[13]. Millions of people are delivering their ideas on social media on various products, services, and events. Along with that, social media also consists of billions of short informal texts that may include SMS, tweets, messages, emails reviews, etc. [10], [14]. This

The associate editor coordinating the review of this manuscript and approving it for publication was Nikhil Padhi.

scenario has brought light upon the topic for researchers to look deep into Sentiment Analysis, Opinion Mining, and Review Analysis because these reviews are potent on any business's survival and downfall. For this reason, it is essential to detect their genuineness. As the popularity of the social web increases, multiple users will keep on spreading various kinds of content almost which lacks any trustworthy external source implying that there is no way of authenticating the content being posted [3], [15], [16]. In the business section, this phenomenon affects an individual consumer and corrupts the confidence of a purchaser in online shopping. Identifying indicators of these fraudulent reviews based on the fraudster's behavior is also an essential task. Due to this, a few scholars have utilized the techniques of Data Mining and Natural Language Processing (NLP) [8], [16], [17] and other techniques such as data cleansing and database query processing to deal with raw data. However, these techniques did not efficiently solve the spam reviews problem. Lately, the reviewers have given plenty of new reviews every day. In this manner, information cleaning and repair will prompt flood in high business activity costs. As the genuineness cannot be identified, it will not be in our interest to approve the database query process that filters those spam. Given the extensive use of social media, intense competition arises in which there is a vital role of consumer reviews which has a great impact on the online marketplace [8], [11], [18]. For improved decision-making, people and organizations need to improve decision-making before purchasing any product [9], [19], [20]. Writing fraudulent comments is mostly done by professionals the establishments hire. These professionals are paid for which they post negative and positive comments on products or brands that are a major help in uplifting or defaming a targeted business [3]. However, these actions of a user could also end up being only a coincidence. One of the principal issues we are confronting today is detecting fake reviews and the extraction of genuine emotion in an opinion. According to American research, 80% of purchasing behavior depends on product feedback. The problem is to determine if the feedback given is genuine or fraudulent. A supervised learning technique is proposed by initially studying the nature of the dataset. We did a thorough analysis of different types of approaches that are working in the same domain. Furthermore, we proposed a technique that shows more remarkable results than state-of-the-art methodologies. Fake reviews are the most pressing issue in the present era. It is one of the most intense topics because it impacts the business world considerably. The gain and loss of businesses partially depend on the feedback, especially in the e-commerce domain. Therefore, it is vital to determine their authenticity by using Machine Learning techniques such as K- Nearest Neighbor, Support Vector Machine, and Logistic regression (SKL).In a recent study,mohawesh et.al. [21] presented a survey of existing models for fake reviews detection. According to this survey, SKL algorithms outperform the accuracy for the proposed problem. The Naive Bayes algorithm is one of the best classification algorithms of machine learning. However, the accuracy of the Naive Bayes algorithm for the detection of fake reviews is slightly less than SKL algorithms [21]. The proposed system includes the following modules;

1. Bi-gram language model
2. Parts-of-Speech tagging
3. Sentiment Analysis
4. Length of review and word count
5. Relationship word count
6. Machine Learning based text classification.

Fake reviews detection techniques are widely used in the e-commerce domain, which plays an essential role in our economy as they can easily uplift or defame a product company or service. Since the purchase decision is firmly motivated by the reviews or ratings, the study shows that work has been concluded in detecting these fraudulent reviews, but spammers' demeanor is constantly developing. Spammers have been discreetly designing these fake reviews to camouflage their malevolent intentions. Many businesses appoint professionals who write inappropriate positive and negative reviews for financial gains. These are fabricated comments that these professionals intentionally write for the sake of seeming authentic. Fake reviews have a powerful impact as they directly influence customers' decision-making power. Relying on the feedback, customers either reject the product or decide to buy the product. Fake reviews are fictitious comments that are either machine-generated or user-generated. Both spams are challenging to identify. In the ongoing years, the use of e-commerce has increased drastically. There have been chances of fraudulent comments that play an essential role in defaming or uplifting a business. Due to the intense competition between organizations, it has become more sophisticated, and thus, many of them use the wrong approach to receive potential profit. Reviews on a product play a part in consumer decisions, and they build confidence in that particular product. However, they cannot be sure about the fallacy of these reviews. Fake reviews can either be deceptive or destructive. Destructive spams are easier to identify by a typical customer since they are non-review and contain ads and messages unrelated to the product. The latter, however, may contain sentimental reviews that may be positive or negative and, thus, problematic. Deceptive review, However, considering the deceptiveness of these reviews, these fake reviews are being used to advance a business or tattle and harm the repute of businesses that are in great competition. The existence of such reviews is crucial for the customer and the business. This concept is known as ''Opinion Mining.'' To bring out the public's mood, Natural Language Processing is used regarding a specific product, service, or company. Untruthful Opinions are negative opinions given to damage the company's repute specifically or promote a business undeservingly. Likewise, positive opinions are given for an organization to gain fame inappropriately. Brand-specific reviews primarily target brands, get negative or positive reviews. Advertisements and irrelevant reviews have no meaning and compromise of no opinion at all. Since the purchase decision is firmly motivated by

the reviews or ratings, the study shows that work has been concluded in detecting these fraudulent reviews, but spammers' demeanor is constantly developing. Spammers have been discreetly designing these fake reviews to camouflage their malevolent intentions. Many businesses appoint professionals who write inappropriate positive and negative reviews for financial gains. These are fabricated comments that these professionals intentionally write for the sake of seeming authentic. Fake reviews have a powerful impact as it directly influences the customer's decision-making power. Relying on the feedback, customers either reject the product or decide to buy the product. The product's price is a significant factor for a consumer, but feedback or reviews on those products are also considered seriously when purchasing something online. Building a trust factor is essential because most people rely on feedback to make a purchase online.

The following aspects sum up the novelty of this research: First, feature selection is based on a multi-level feature extraction system. Besides the normal Natural Language Processing (NLP) on the corpus to extract and feed features to the classifiers, this research proposed several feature engineering techniques to extract various behavior of the reviewer himself and reviews. Further, behavioral features were also extracted for feature engineering. Behavioral features represent the statistical significance of a user's review. They may not directly contribute to the classification accuracy; instead, they have linked to the reviewer himself. For example, review time, writing style, use of punctuation, verb/noun count in a review, and relationship words. All these features contributed to the results' overall classification accuracy and authenticity. Secondly, we tried to get the best fit training and testing dataset samples to get the best classifiers results.

The paper is arranged in sections as sections II covers literature review, the proposed methodology is explained in section III, section IV comprises design and implementation including results. A conclusion has been drawn in section V.

II. LITERATURE REVIEW

Spam reviews are fictitious comments that are either machine-generated or user-generated. Both spams are challenging to identify. In recent years, with the increasing use of e-commerce online, there have been chances of fraudulent comments that play an essential role in defaming or uplifting a business. Due to the intense competition between organizations, it has become more sophisticated, and thus, many of them use the wrong approach to receive potential profit. Reviews on a product play a part in consumer decisions and build confidence in that particular product. However, they cannot be sure about the fallacy of these reviews. Spams can either be deceptive or destructive. Destructive spams are easier to identify by a typical customer since they are non-review and contain unrelated ads and messages unrelated to the product. The latter, however, may contain sentimental reviews that may be positive or negative and, thus, problematic. The existence of such reviews is crucial for the customer and the business. This concept, in other words, is also called "Opinion Mining." It is a technique in Natural Language Processing to figure out the public's mood regarding a specific product, service, or company. However, considering the deceptiveness of these reviews, these fake reviews are being used to promote a business or spread rumors and harm the reputation of competing businesses. Since the purchase decision is firmly motivated by the reviews or ratings, a study shows that work has been concluded in detecting these fraudulent reviews, but spammers' demeanor is constantly developing. Spammers have been discreetly designing these fake reviews to camouflage their malevolent intentions. Many businesses appoint professionals to write inappropriate positive and negative reviews for financial gains. These are fabricated reviews that are intentionally written to seem authentic. Deceptive spam review is harmful to the repute of any product as it misleads the customer to make decisions. Somayeh *et al.* [19] came up with a lexical and syntactical feature technique using machine learning classifiers to detect spam or ham. The features include n-gram, Part of speech (POS) tagging, and LIWC (Linguistic Inquiry and Word Count). They took deceptive reviews from Amazon.com and truthful reviews from TripAdvisor.com. Their results showed 81% accuracy with Naïve Bayes (NB) classification algorithm and 70% with Sequential Minimal Optimization (SMO) using lexical features. Moreover, using syntactic features gave 76% and 69% accuracy using the same classifiers. At the same time, their combination gave 84% and 74% with NB and SMO. However, the results did not exceed 85%—furthermore, Rajamohana *et al.* [22] proposed a methodology for detecting opinion spam using features detection. They proposed an approach that deals with selecting subset features from many feature sets for the classifier to separate spam or ham. The two approaches utilized are cuckoo search, and hybrid improved binary particle swarm optimization (iBPSO), Naïve Bayes, and KNN classifiers that are helping in the classification process. These two approaches have been compared, and a hybrid search achieved a comparatively higher accuracy measure. However, this approach is solely dependent on feature selection. Moreover, Catal and Guldan [23] came up with supervised and unsupervised techniques to know by sight the spam review. There is a significant chance that spam reviewer is responsible for the content pollution in social media as many users have multiple login IDs. The researchers tackled that problem and utilized the most productive feature sets to structure their model. Semantic analysis is also unified in the detection process. In addition, some standard classifiers are applied on labeled datasets, and for unlabeled datasets, clustering is used after desired attributes. They worked with both labeled and unlabeled data along with a unigram model and achieved 86% results. Ott *et al.* [24] proposed a model to identify fraudulent consumer reviews using multiple classifiers in online shopping. The selected classification techniques were majority voted libLinear, libSVM, minimal sequential optimization, random forest, and J48. Then the evaluation was compared with other models,

SVM technique with 5-fold cross-validation to get 86% was accuracy. Rout *et al.* [3] explained that how semi-supervised classifiers are used to detect online spam reviews using a dataset of hotel reviews. Dissimilar to other different kinds of spam [1], [3] it is demanding to recognize an unreal opinion as it is needed to understand the contextual meaning to know the nature of the review. Supervised learning is conventionally used to detect fake reviews, but it also has some restrictions, such as assurance of the quality of reviews in the training dataset. Secondly, to train the classifier, it can be challenging to obtain the data because of the diverse nature of the online reviews. The limitations mentioned above can be overcome using a semi-supervised learning approach by unifying three new dimensions to the domain of the feature as in POS feature, Linguistic and Word Count Feature, and Sentimental Content features to get more significant results. A dataset of both positive and negative reviews has been used. They, however, achieved an 83% f-score. He *et al.* [25] introduced the rumors model and applied the text mining technique, and extracted three notable characteristics of the content of reviews such as noun/verb ratio, important attribute word, and a specific quantifier. TripAdvisor dataset was used, and results showed that the unique vocabulary, specific quantifiers, and nouns it contains, the more valuable and truthful the review is. Moreover, the results showed 71.4% F-measure, 60% accuracy, 86% recall, and a fake evaluation value of 0.016952338. Meaning, higher the fake evaluation value, the more fake a review is. Deceptive opinions are more fictitious but sound real. People are hired by many businesses to write unjustified reviews about the products which are undistinguishable by the people. Therefore, Ott *et al.* [24] performed a test that gave the accuracy of 57.33% of three human judges, which made this research even more valid, significant, and pithy. However, it is hard to define the semantic perspective from the data. Significant donations of the paper are; firstly, to understand the semantic better, a document level review is represented. Secondly, multiple syntax features are used to make a feature combination to improve performance. Thirdly, domain-independent and domain migration experiments verify the SWNN and feature combination performance. Further, in the domain of neural networks, Goswami *et al.* [26] proposed a feature set by observing the user's social interaction behavior to recognize reviewer hoaxes. They used a neural network to analyze the feature set and compare it with other contemporary feature set in detecting spam. Features include the number of friends, followers, and number of times a user has provided enough room to form a relationship between opinion spam and social interaction behavior. Aside from neural networks, most scholars focused on supervised learning techniques. Therefore, Brar and Sharma [16] proposed an approach that is used to analyze the review and reviewer-centric feature to detect fake reviews using the supervised learning technique. It provided comparatively better results than completely unsupervised learning techniques, mostly graph-based methods. A publically available large-scale and standard data set from a review site Yelp.com [27] has been considered here and has given more significant results. Furthermore, in the supervised learning domain, Elmurngi and Gherbi [20] analyzed the online reviews for movies using Sentiment Analysis (SA) methods and text classification for the sake of recognizing fake reviews. The scholars presented the classification of the movies review as positive or negative by using machine learning (ML) methods. The comparison between five individual ML classifiers, Naïve Bayes (NB), SVM, KNN- IBK, K*, and DT-J48, for sentiment analysis is made using two datasets that include movie review datasets V1.0 and movie review dataset V2.0. Some researchers also focused on different factors in determining fake reviews, such as Arjun Mukharjee *et al.* [5] pay attention to fake reviewers groups instead of individual reviews; therefore, they came up with the frequent itemset mining method to identify the groups. Furthermore, they built a labeled dataset of the reviewers' group. The results showed that their methodology outperformed the standard classification techniques using the Amazon dataset. In order to determine negative reviews on crowdsourcing platforms, Parisa *et al.* [18] observed the behavior of the reviews on these sites and observed the behavior of the reviews given. They indicated clues on the detection process of such manipulating reviews that are fake yet hiding in plain sight. However, this approach is risky because it relies on observations that may or may not be accurate. On the other hand, Shebuti Rayana *et al.* [4] mainly focused on two methodologies, SP Eagle and Fraud Eagle, and did a comparison using utilized clues from metadata (timestamps, text, and rating) and relational data (networks) and created a model for the detection of suspicious behavior, products, and the users by using Yelp.com [27] dataset. Moreover, they derived SP Eagle light from SP Eagle, which is more efficient in computation and it utilizes a minimum set of feature reviews for efficient computations. The primary purpose was to bridge the relational data and metadata to improve the track-down process. Atefeh *et al.* [28] advised a robust spam review detection system to investigate suspicious time intervals of the online reviewers using time series by pattern recognition technique where the results show it to be a better, easy, and more straightforward approach as it gives an F-score of 86% as compared to others [28]. Komal and Sumit [29] described opinion spam and portrayed how it is a genuine concern these days. Since fuzzy logic deals with real-life uncertainties, a novel solution based on fuzzy modeling is proposed. Four fuzzy logic input linguistic variables are considered, and the spammer group's suspicious level is termed as Ultra, Mega, Immense, Highly, Moderate, Slightly, and feebly. A novel algorithm has been used that utilizes 81 rules of fuzzy logic and fuzzy Ranking Evaluation Algorithm (FREA) to refract the extent of the spam's suspiciousness. The datasets are used to satisfy the 3 V's of Big Data; hence Hadoop is used for the storage and analysis. The proposed algorithm further demonstrated using sample review's data sets and amazon data sets, achieving an accuracy of 80.77%. S.P.Rajamohana *et al.* [22] came up with a feature selection technique that was effective. It is called a cuckoo search in junction with harmony search.

In contrast, Naïve Bayes is used to categorizing spam or ham. Evolutionary algorithms are used for feature selection, which can handle the high spatiality of the feature removing irrelevant, noisy features and considering the excellent feature selection to increase the processing rate and predictive accuracy. Yuming lin *et al.* [17] dealt with the detection of fake reviews in review sequence. They observed the characteristics of fake reviews that depend on the contents of the reviewer's behavior. They also introduced six times more sensitive features that include modeling the review content, the similarity of reviews on content, the similarity of reviews on a product and other products, modeling the frequency of the reviews made by the reviewers, repeatability measures, and frequency of the review. As a result, the identification of spam reviews was orderly and in high precision. Muhammad *et al.* [15] investigated the performance of the rule-based machine learning technique, which is a learning classifier system (LCS), in semantic analysis of Twitter messages, movie reviews, and spam detection from SMS and email data sets. The results showed that the proposed methods smoothed the learning process and gave better results in the experiments. Furthermore, Hamza Al Najadah *et al.* [2] introduced a bagging-based approach to balance the imbalanced datasets than using supervised learning they have done the classification. Using datasets from Amazon Turk from the deceptive opinion spam corpus volume 1.4, their results showed better precision, recall, and accuracy than standard classifiers. Furthermore, Fusilier *et al.* [1] came up with the PU learning technique, which is a semi-supervised classification technique to cater to both types of deceptive reviews, positive and negative. The proposed methodology selecting negative features was a bit unprogressive, but the results showed an improvement of 8.2% and 1.6% over the original model. Most of the best researchers used supervised learning techniques alongside other different approaches and determined f-score almost close to 90%. However, our proposed technique was better than state-of-the-art techniques and showed better accuracy.

III. PROPOSED METHODOLOGY AND EXPERIMENT SETUP

We proposed a support vector machine, K-Nearest Neighbor and Linear Regression (SKL) based algorithm for fake reviews detection in the e-commerce industry. To fulfill our objective, we observed a dataset on hotel reviews and applied machine learning techniques and text classification methods to detect reviews that are not genuinely made. Fig 1 shows the steps of the proposed methodology.

A. PREPROCESSING

Preprocessing data phase includes the filtering process. It represents the part where we get rid of the text's less valuable parts, such as punctuation symbols. Punctuation marks such as,''.!?;.. etc. are eliminated because it lowers the overall accuracy of the classification process. Their removal results in better output by the algorithm used. In order to complete this process, Natural Language ToolKit (NLTK) package is used. After successfully removing the punctuation words, word count is calculated. Selecting variables or identifying attributes to construct an efficient model is called feature selection. The objective of this process is to achieve a higher level of accuracy. In our proposed method, feature selection is based on the following parameters

1. Length count
2. Bigram Type
3. Relationship words
4. Sentiment word count
5. Noun, Verb count

Firstly, the total length of review is calculated, and then by using the bigram probability model, the probability of the next coming word is calculated. This is also called the Markov model, where you can define the probability of the next coming word without looking at it in the complete document. Some words describe the relationship just like husband, wife, sister, niece, etc. In-text classification analysis, we called these words relationship words. SKL selects features by considering relationship words. A list of relationship words is created and used for SKL based proposed solution.Feature selection also depends on the sentiment of word count, whether it is a positive or negative word in a review. Corpora or bag of the word is created with positive and negative words. In order to figure out the sentiment of a review word, those words match the pre-calculated corpora (positive or negative) of the given dataset. For this purpose, we have created positive and negative corpora, which contain approximately 2006 positive words and 4783 negative words. Part of speech (POS) tagging marks the corresponding word or part of speech in the given sentence. In our proposed model, NLTK is used to tokenize the sentence, and then by using POS, they are tagged as noun-verb or adverb, etc. In the proposed model, Noun and verb counts are also calculated as part of the feature selection process. We split the dataset in an 80-20 ratio for training and testing samples, respectively. For more refined results, 10 fold cross-validation is done, leading to 95% and 89.03% overall classification accuracy on the Yelp and the TripAdvisor dataset.

TABLE 1. Relationship word corpora.

1. wife
2. husband
3. children
4. child
5. son
6. daughter
7. aunt
8. uncle
9. nephew
10. niece
11. sister
12. brother
13. grandfather
14. grandmother
15. father
16. mother

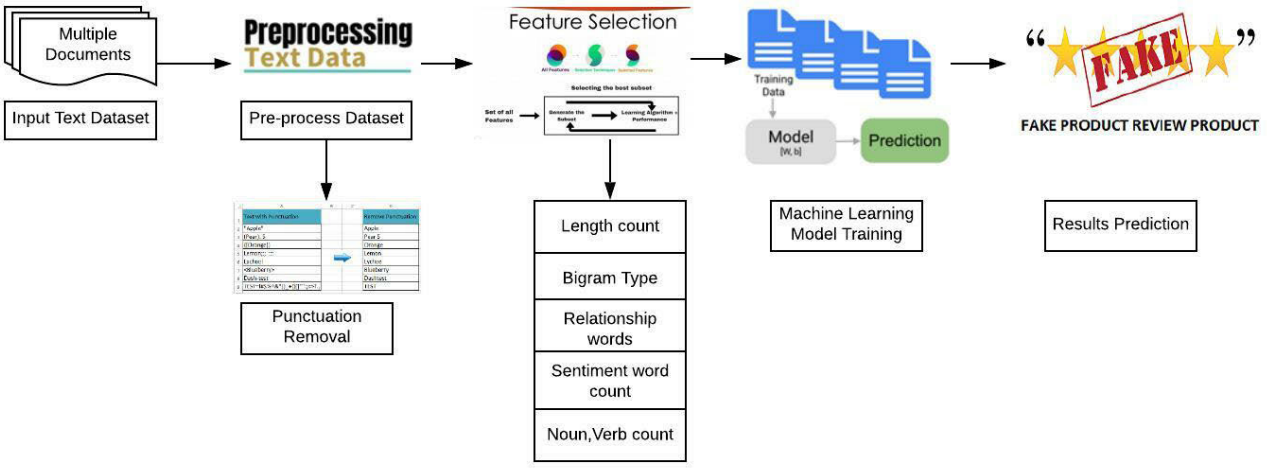

FIGURE 1. Block Diagram of proposed SKL base fake reviews detection methodology.

TABLE 2. Summary of the yelp dataset.

Total number of reviews	1900 reviews
Number of fake review	950 reviews
Number of real reviews	950 reviews
Total word count in dataset	49060 words
Positive word corpora	2006 words
Negative word corpora	4783 words
The maximum review length	452 words
The minimum review length	15 words
The average review length	119.7 words

TABLE 3. Datasets details.

Dataset	Yelp	Trip advisor
Data Construction Method	Filtering Algorithm	Amazon Mechanical Turk (AMT)
Total Reviews	1900	800
Domain	Restaurants and Hotels	Hotels

B. CLASSIFICATION ALGORITHMS

Classifying data into two or more than 2 classes/Labels is called classification. Machine Learning comes with many classification algorithms. In a recent study [21], they came up with a survey report and concluded that machine learning algorithms perform well on medium-sized datasets as compared to Artificial Neural Network (ANN) and deep learning models, which are a good fit for large size datasets. Feature engineering is another reason ANN and deep learning models do not show better results for fake review detection. Feature selection is an integral part of machine learning models training and plays a vital role in getting better classification results. However, in ANN, there is not a straightforward process for feature selection. ANN converts the word vector of reviews into a matrix. This matrix is fed to the convolutional layer by following different filters and forward results to the pooling layer. ANN is a complete "black box" without any information about feature selection. Therefore,

Algorithm 1 Feature Selection for SKL Based Model.

Input: True labeled training data t-train.txt as T
False labeled training data f-train.txt as F
Output: Selected feature as SF.txt
Read T
Remove Punctuation
Generate genuineReview list as G
Read F
Remove Punctuation
Generate fakeReview list as G′
Merge G and G′ as U
for *each review as x in U*, $\forall x \in U$ **do**
List.append(*GetLengthOfReview*)
List.append(*GetNumberofBigramTypes*)
List.append(*GetNumberofRelationshipWords*)
List.append(*GetSentimentWordCount*)
List.append(*GetNumberofPronounsAndVerbs*)
end for
$train_f eatures.append(List)$
if $rev[\text{"}sentiment\text{"}] == \text{"}True\text{"}$ **then**
$train_f eatures.append(1)$
else
$train_f eatures.append(0)$
end if
return SF.txt

the classification algorithms we utilized are Support Vector Machine (SVM) with linear SVC (Support Vector Classifier) kernel to predict that either given product review is fake or genuine. SVM is a pattern detection model in supervised learning, also associated with learning algorithms used for classification and prediction. It draws a decision boundary, also known as a hyperplane, near the extreme points of the dataset after identifying those extreme points inside the dataset. The K-Nearest Neighbors (K = 5) is also used for pattern recognition and classification. It is a straightforward

algorithm. Its performance depends on many factors, such as the k parameter, an acceptable measure distance, and a majority voting scheme. In statistics, logistic regression (LR) is also a part of machine learning. It is a technique that is used in binary classification. It uses the logistic function for prediction purposes. We further used logistic regression to predict the nature of a review. All three classification algorithms are part of fake reviews detection.

C. DETECTION PROCESS

After completing the training phase, the dataset will test the model to predict the output. Model is being trained by SVM, KNN, and Logistic regression. The comparison table shows which algorithm outperforms for the selected process.

D. EXPERIMENTATION DESIGN / DETAILS

We based our experiment on determining fake reviews that play an essential role in the progress of online businesses. Dataset used in our proposed research is self extracted using filtering method from Yelp.com [27]. Fake reviews extracted from the yelp website are more realistic than deceptive datasets representing semi-real data. Moreover, fake review detection is more challenging with realistic datasets with overlapping between legitimate and fake review data. [21].

[24] developed a dataset for fake reviews detection, they bypassed reviews having a length less than 150 characters. We assumed these facts from literature and created a considerably more extensive dataset, including mixed length reviews that vary from 452 words (1808 characters) to 15 words (60 characters). Yelp dataset [27] is an imbalanced dataset. Moreover, this dataset is biased to positive reviews at the expense of detecting negative fake reviews. We developed our self-extracted dataset from Yelp's data file to this effect. We used the filtering method to extract a subset of Yelp's dataset and validated it with human judgments.

Two factors were assessed in the data extraction process—first, mixed reviews. Secondly, biasness of any class (positive/negative). We ensure to take an equal number of positive and negative reviews to avoid imbalance dataset issues.

1) DATASET COLLECTION

We collected a dataset of 20 hotel reviews from Yelp.com [27]. It includes 1900 reviews in which 950 are harmful, and the rest of the 950 are positive reviews, which makes a balanced dataset. For building a model with good generalization performance, the best data splitting strategy is essential for every classification model, which is crucial for model validation. For the performance assessment of a model, it is a practice to divide the available dataset into two subsets in the ratio of 4:1 for the training dataset and testing dataset. However, we may need to readjust this ratio most of the time to get better performance. Moreover, the ratio may vary from one classification algorithm to the other [30]. We split the dataset in an 80-20 ratio, of which 80% is the training set, and 20% is the test set. To assess the performance of the best suitable model for the proposed problem, we then split our data into 75-25 and 85-15 ratio split.

One of the challenging issues in fake reviews detection is the availability of labeled datasets. It is observed that most of the available datasets are constructed based on a crowdsourcing framework. Ott *et al.* [24] developed an opinion spam dataset with gold standard deceptive opinion. We extend our experimentation to another gold standard dataset proposed by ott *et al.* [24]. This self-generated dataset consists of 800 reviews from TripAdvisor. They collect positive reviews from the 20 most popular hotels from the Chicago area, including 5-star truthful reviews. while deceptive opinions are collected from the same 20 hotels using Amazon Mechanical Turk (AMT).

E. STATISTICAL ANALYSIS

A statistical test was run on the experimental data to compare the different learning models employed in this study. We have chosen the Friedman test [31] in particular because this test was proposed to compare several classifiers approaches on a variety of datasets. The Friedman test is based on a ranking of each classification method in each dataset, with the best algorithm receiving rank 1, the second-best algorithm receiving rank 2, and so on. In this rank, ties are broken by taking the average of their ranks.

We compare k algorithms on N distributions of the dataset, with r_i denoting the i^{th} algorithm's average order value. If we ignore the halved value for the time being, r_i follows the normal distribution, with mean and variance of $(k+1)/2$ and $(k^2-1)/12$, respectively. After that, the Friedman statistic with k-1 degrees of freedom is as follows:

$$\tau_{x^2} = \frac{k-1}{k} \cdot \frac{12N}{k^2-1} \sum_{i=1}^{k} \left(r_i - \frac{k+1}{2} \right)^2 \tag{1}$$

$$= \frac{12N}{k(k+1)} \left(\sum_{i=1}^{k} r_i - \frac{k(k+1)^2}{4} \right) \tag{2}$$

Nonetheless, it is demonstrated that there is a more relevant statistic with $k-1$ and $(k-1)(N-1)$ degrees of freedom that is distributed according to the F-distribution [31]. The Friedman F is a statistic that is expressed as follows:

$$\tau_F = \frac{(N-1)\tau_{x^2}}{N(k-1) - \tau_{x^2}} \tag{3}$$

If the Friedman test's null hypothesis is rejected, posthoc tests can be used to supplement the statistical analysis [32]. The Nemenyi tests were used in this study to get insight into the differences between the tested classifiers. All classifiers are compared to one another in the Nemenyi test [31]. If two classifiers' rankings differ by at least the crucial difference in this method, their performance is significantly different. The following equation calculates the critical difference:

$$CD = q_\alpha \sqrt{\frac{k(k+1)}{6N}} \tag{4}$$

TABLE 4. Performance metrics for SVM, KNN, and LR classification algorithms.

Dataset Split	Classification Algorithm	Accuracy	Precision	Recall	F-score
75-25	SVM	0.94	0.89	0.95	0.92
	KNN	0.83	0.78	0.81	0.79
	LR	0.79	0.75	0.82	0.78
80-20	SVM	0.95	0.90	1	0.94
	KNN	0.85	0.80	0.80	0.80
	LR	0.85	0.81	0.85	0.85
85-15	SVM	0.91	0.90	0.96	0.92
	KNN	0.82	0.79	0.84	0.81
	LR	0.81	0.76	0.81	0.78

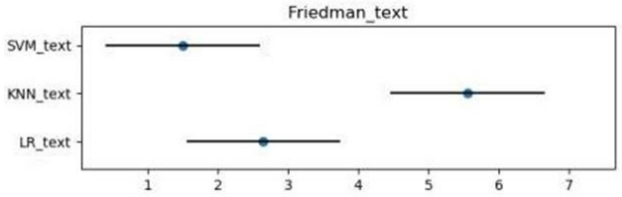

FIGURE 2. Statistical Analysis based on Friedman test chart.The average order value is on the horizontal axis, while each algorithm is on the vertical axis. Each algorithm's average order value is shown by a dot.

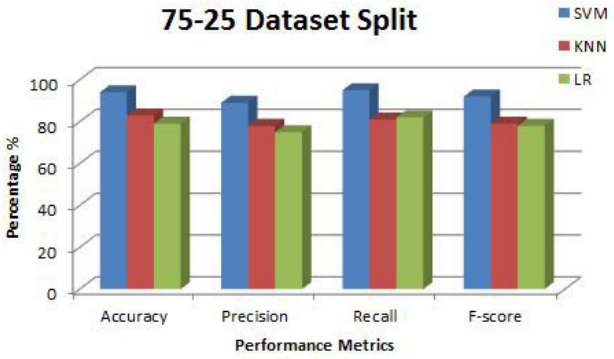

FIGURE 3. 75-25% Dataset Split.

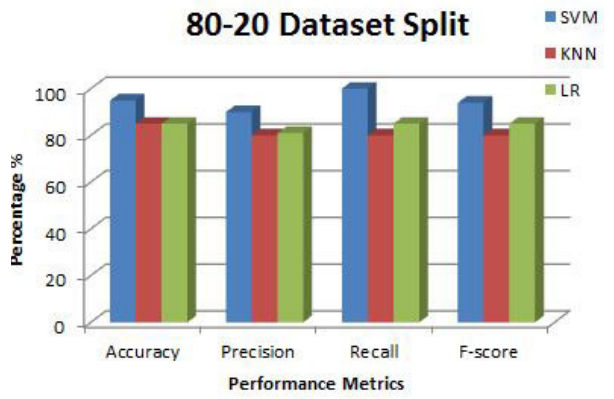

FIGURE 4. 80-20% Dataset Split.

The following parameters shows how the tests were computed on the text data set. The ranks have been obtained in respect to the Friedman test.The α value is set to 0.05 for all calculations [31]. $\tau_{x^2_{text}} = 40.5$, $\tau_{F_{text}} = 18.6$, the critical value of F distribution $F(k-1,(k-1)(N-1)) = 2.27$, so $\tau_{F_{text}} > F(6.54)$, the negative hypothesis is rejected, that is, not all classifiers have similar performance, so they can be tested later. After that, run the Nemenyi test. According to the DemZar J [31] table query, $q_\alpha = 2.272$, and based on these data, calculate the critical value $CD_{text} = 2.19$.

The Friedman test chart is generated as shown in Figure based on the experimental results. Nemenyi test reveals that SVM is considerably different from KNN and LR classifiers based on the rank obtained in the Friedman test. The performance of SVM is significantly better than that of LR and KNN.

IV. FINDINGS AND DISCUSSION

Results are calculated using the following classification evaluation metrics: precision, recall, f-score, and accuracy. Table 4 shows the classification results of different classifiers on different dataset splits. As described previously, we check classification results on 75% training set and 25% test set data distribution by using 10 fold cross-validation. Further, we check the results on the 80-20 dataset split and 85-15 dataset split. As shown in Table 4 and also in Fig3, Fig.4, and Fig 5, Support Vector Machine (SVM) outperforms as compared to K Nearest Neighbor (KNN) and Logistic Regression (LR). The best dataset splits in this experimentation is 80-20. The results show that SVM has the highest recall, which means that the prediction process is efficient in SVM. Further, KNN and Logistic regression show a straight line of 90 percent, indicating these both classifiers have the same recall value that is lesser than SVM.

TABLE 5. Performance Comparison of the proposed SKL model and the other existing models.

Classification Algorithms	Accuracy
SKL (Proposed Model)	95%
SVM, NB [20]	81.35%,79.7%
NB, KNN [33]	82.3%
SVM [23]	93%
SVM, NB, DT [34]	86.25%

In table 5 we compared our results with the state-of-the-art methodologies using a similar dataset. These approaches used different classifiers with similar techniques, and as a result

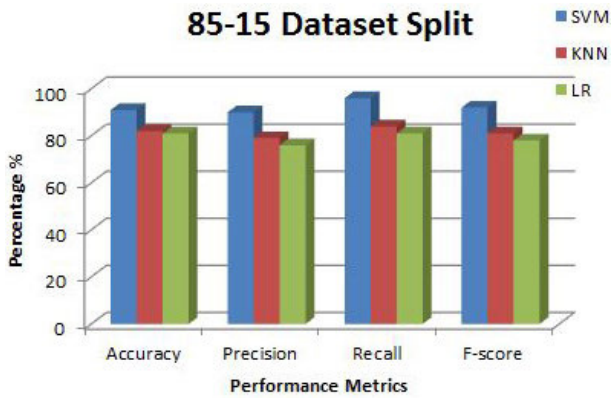

FIGURE 5. 85-15% Dataset Split.

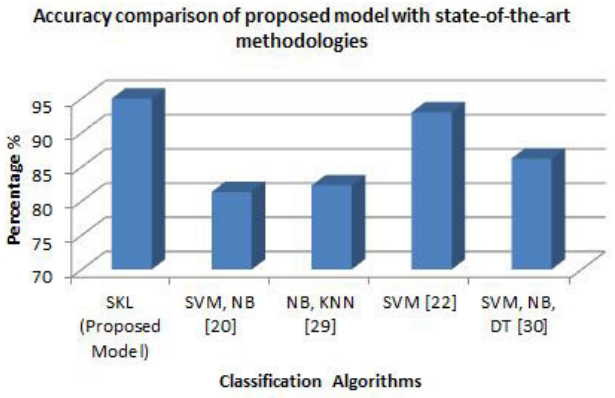

FIGURE 6. Comparison between the performance of the proposed SKL model and the other existing models.

TABLE 6. Performance Comparison of the proposed SKL model and the other existing models on similar dataset.

Classification Algorithms	Dataset	Accuracy
SKL (Proposed Model)	Yelp Dataset	95%
SVM [5]	Yelp Dataset	86%
SVM. NB, RF [35]	Yelp Dataset	79%,52%,84%

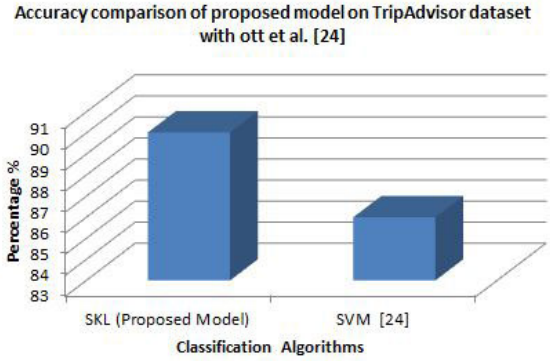

FIGURE 7. Accuracy comparison of proposed model on TripAdvisor dataset with ott *et al.* [24].

of this experiment, our proposed model reached 95% accuracy on Yelp dataset compared to existing models. In Fig 6, a comparison of accuracy results has been shown. The preprocessing step of the proposed model, along with the most suitable machine learning algorithm SVM in this research, outperforms other models. In order to achieve robustness of our proposed methodology, We extended experimentation on the TripAdvisor dataset [24]. Since SVM outperformed the previous experiment, and 80-20% dataset split is figured out as the best split for training and testing, we use SVM with the proposed feature engineering method on the TripAdvisor dataset. we get better accuracy results of 89.03% compared to ott *et al.*[24] results of 86% on the same dataset as shown in Table 6 and Fig 7.Although the accuracy difference is not promising on the TripAdvisor dataset, it somehow proves the effectiveness of the proposed methodology.

V. CONCLUSION AND FUTURE WORK

Throughout this research, it has been observed that fake reviews are indeed hard to tackle. Many studies have been working on this topic, but no study has given a one hundred percent result. Even in the present era, many loopholes are not being addressed. We proposed a methodology using machine learning-based text classification that helped determine whether the given comments on a particular product/service are real or fake. Our SKL technique proved to be more robust than already existing methodologies in the same field and proved to be more accurate. The results prove that SKL based fake review provides 95% on Yelp dataset and 89.03% accuracy on TripAdvisor dataset compared to other state-of-the-art techniques. Since most of the researchers mainly focused on a complete supervised learning process. Therefore, in the future, we would like to study Positive-unlabeled (PU) learning techniques in depth, which is a semi-supervised learning approach.

ACKNOWLEDGMENT

The authors, therefore, gratefully acknowledge the DSR technical and financial support.

REFERENCES

[1] D. H. Fusilier, M. M.-Y. Gómez, P. Rosso, and R. G. Cabrera, "Detecting positive and negative deceptive opinions using pu-learning," *Inf. Process. Manage.*, vol. 51, no. 4, pp. 433–443, 2015.

[2] H. A. Najada and X. Zhu, "ISRD: Spam review detection with imbalanced data distributions," in *Proc. IEEE 15th Int. Conf. Inf. Reuse Integr. (IEEE IRI)*, Aug. 2014, pp. 553–560.

[3] J. K. Rout, S. Singh, S. K. Jena, and S. Bakshi, "Deceptive review detection using labeled and unlabeled data," *Multimedia Tools Appl.*, vol. 76, no. 3, pp. 3187–3211, Feb. 2017.

[4] 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*, Aug. 2015, pp. 985–994.

[5] 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.

[6] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What yelp fake review filter might be doing," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 7, 2013.

[7] J. K. Rout, A. Dalmia, K.-K. R. Choo, S. Bakshi, and S. K. Jena, "Revisiting semi-supervised learning for online deceptive review detection," *IEEE Access*, vol. 5, pp. 1319–1327, 2017.

[8] W. Etaiwi and G. Naymat, "The impact of applying different preprocessing steps on review spam detection," *Proc. Comput. Sci.*, vol. 113, pp. 273–279, Jan. 2017.
[9] S. Choi, A. S. Mattila, H. B. Van Hoof, and D. Quadri-Felitti, "The role of power and incentives in inducing fake reviews in the tourism industry," *J. Travel Res.*, vol. 56, no. 8, pp. 975–987, Nov. 2017.
[10] L. J. Sheela, "A review of sentiment analysis in Twitter data using Hadoop," *Int. J. Database Theory Appl.*, vol. 9, no. 1, pp. 77–86, Jan. 2016.
[11] E. Aydogan and M. A. Akcayol, "A comprehensive survey for sentiment analysis tasks using machine learning techniques," in *Proc. Int. Symp. Innov. Intell. Syst. Appl. (INISTA)*, Aug. 2016, pp. 1–7.
[12] L. Zhang, Y. Yuan, Z. Wu, and J. Cao, "Semi-SGD: Semi-supervised learning based spammer group detection in product reviews," in *Proc. 5th Int. Conf. Adv. Cloud Big Data (CBD)*, Aug. 2017, pp. 368–373.
[13] B. Liu, "Sentiment analysis and opinion mining," *Synthesis Lectures Hum. Lang. Technol.*, vol. 5, no. 1, pp. 1–167, 2012.
[14] N. J. Conroy, V. L. Rubin, and Y. Chen, "Automatic deception detection: Methods for finding fake news," in *Proc. 78th ASIST Annu. Meeting, Inf. Sci. Impact, Res. Community*, vol. 52, no. 1, 2015, pp. 1–4.
[15] M. H. Arif, J. Li, M. Iqbal, and K. Liu, "Sentiment analysis and spam detection in short informal text using learning classifier systems," *Soft Comput.*, vol. 22, no. 21, pp. 7281–7291, Nov. 2018.
[16] G. S. Brar and A. Sharma, "Sentiment analysis of movie review using supervised machine learning techniques," *Int. J. Appl. Eng. Res.*, vol. 13, no. 16, pp. 12788–12791, 2018.
[17] Y. Lin, T. Zhu, H. Wu, J. Zhang, X. Wang, and A. Zhou, "Towards online anti-opinion spam: Spotting fake reviews from the review sequence," in *Proc. IEEE/ACM Int. Conf. Adv. Social Netw. Anal. Mining (ASONAM)*, Aug. 2014, pp. 261–264.
[18] P. Kaghazgaran, J. Caverlee, and M. Alfifi, "Behavioral analysis of review fraud: Linking malicious crowdsourcing to Amazon and beyond," in *Proc. Int. AAAI Conf. Web Social Media*, vol. 11, 2017.
[19] S. Shojaee, M. A. A. Murad, A. B. Azman, N. M. Sharef, and S. Nadali, "Detecting deceptive reviews using lexical and syntactic features," in *Proc. 13th Int. Conf. Intellient Syst. Design Appl.*, Dec. 2013, pp. 53–58.
[20] E. I. Elmurngi and A. Gherbi, "Unfair reviews detection on Amazon reviews using sentiment analysis with supervised learning techniques," *J. Comput. Sci.*, vol. 14, no. 5, pp. 714–726, May 2018.
[21] R. Mohawesh, S. Xu, S. N. Tran, R. Ollington, M. Springer, Y. Jararweh, and S. Maqsood, "Fake reviews detection: A survey," *IEEE Access*, vol. 9, pp. 65771–65802, 2021.
[22] S. P. Rajamohana, K. Umamaheswari, and S. V. Keerthana, "An effective hybrid cuckoo search with harmony search for review spam detection," in *Proc. 3rd Int. Conf. Adv. Electr., Electron., Inf., Commun. Bio-Inform. (AEEICB)*, Feb. 2017, pp. 524–527.
[23] C. Catal and S. Guldan, "Product review management software based on multiple classifiers," *IET Softw.*, vol. 11, no. 3, pp. 89–92, Jun. 2017.
[24] 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.
[25] X. He, X. Gao, Y. Zhang, Z.-H. Zhou, Z.-Y. Liu, B. Fu, F. Hu, and Z. Zhang, "Intelligence science and big data engineering. Big data and machine learning techniques," in *Proc. 5th Int. Conf., (IScIDE)*, vol. 9243, Suzhou, China: Springer, Jun. 2015, pp. 29–42.
[26] K. Goswami, Y. Park, and C. Song, "Impact of reviewer social interaction on online consumer review fraud detection," *J. Big Data*, vol. 4, no. 1, pp. 1–19, Dec. 2017.
[27] *Hotel Reviews Dataset*. Accessed: Jun. 11, 2019. [Online]. Available: https://www.yelp.com/dataset
[28] F. Atefeh and W. Khreich, "A survey of techniques for event detection in Twitter," *Comput. Intell.*, vol. 31, no. 1, pp. 132–164, 2015.
[29] K. Dhingra and S. K. Yadav, "Spam analysis of big reviews dataset using fuzzy ranking evaluation algorithm and Hadoop," *Int. J. Mach. Learn. Cybern.*, vol. 10, no. 8, pp. 2143–2162, Aug. 2019.
[30] A. Krishna *et al.*, "Sentiment analysis of restaurant reviews using machine learning techniques," in *Emerging Research in Electronics, Computer Science and Technology*. Singapore: Springer, 2019, pp. 687–696.
[31] J. Demšar, "Statistical comparisons of classifiers over multiple data sets," *J. Mach. Learn. Res.*, vol. 7, pp. 1–30, Jan. 2006.
[32] 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.
[33] P. K. Sa, M. N. Sahoo, M. Murugappan, Y. Wu, and B. Majhi, "Progress in intelligent computing techniques: Theory, practice, and applications," in *Proc. ICACNI*, vol. 2. Singapore: Springer, 2017, pp. 265–271.
[34] D. Zhang, L. Zhou, J. L. Kehoe, and I. Y. Kilic, "What online reviewer behaviors really matter? Effects of verbal and nonverbal behaviors on detection of fake online reviews," *J. Manage. Inf. Syst.*, vol. 33, no. 2, pp. 456–481, 2016.
[35] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "Fake review detection: Classification and analysis of real and pseudo reviews," Univ. Illinois, Chicago, IL, USA, Tech. Rep. UIC-CS-03-2013, 2013.

HINA TUFAIL received the M.S. degree from the National University of Computer and Emerging Sciences, Lahore, in 2016. She has been working as a Lecturer with the Department of Computer Science, University of Management & Technology, Sialkot Campus, Pakistan, since March 2016. She has been supervising numerous projects in the domain of medical image processing using deep learning architectures. Her research interests include computational intelligence, machine learning, natural language processing, and deep learning.

M. USMAN ASHRAF received the Ph.D. degree in computer science from King Abdulaziz University, Saudi Arabia, in 2018. He was a High-Performance Computing (HPC) Scientist with the HPC Centre, King Abdulaziz University. He is currently an Assistant Professor and the Head of the Department of Computer Science, GC Women University Sialkot, Pakistan. His research interests include exascale computing systems, high-performance computing systems, parallel computing, HPC for deep learning, and location-based services system has appeared in IEEE Access, *IET Software*, the *International Journal of Advanced Research in Computer Science*, the *International Journal of Advanced Computer Science and Applications*, the *International Journal of Information Technology and Computer Science*, the *International Journal of Computer Science and Security*, and several international IEEE/ACM/Springer conferences.

KHALID ALSUBHI received the B.Sc. degree in computer science from King Abdulaziz University (KAU), in 2003, and the M.Math. and Ph.D. degrees in computer science from the University of Waterloo, Waterloo, ON, Canada, in 2009 and 2016, respectively. He is currently an Assistant Professor of computer science at KAU. His research interests include network security and management, cloud computing, and security and privacy of healthcare applications.

HANI MOAITEQ ALJAHDALI was born in Jeddah, Saudi Arabia, in 1983. He received the B.Sc. degree in computer science from King Abdulaziz University, Jeddah, in 2005, and the M.Sc. degree in information technology and the Ph.D. degree in computer science from the University of Glasgow, in 2009 and 2015, respectively. From 2005 to 2007, he worked at Saudi Electricity Company as a Budget and System Analyst. In 2011, he has appointed as a Lecturer at the Department of Information Systems, King Abdulaziz University. He is currently appointed as an Associate Professor at the Faculty of Computing and Information Technology in Rabigh, King Abdulaziz University. His research interests include information security, human–computer interaction, and machine learning.