

Intelligent Interface for Fake Product Review Monitoring and Removal

Ata-Ur-Rehman¹, Nazir M. Danish², Sarfraz M. Tanzeel³, Nasir Usama⁴, Aslam Muhammad⁵, Martinez-Enriquez A. M.⁶, Adrees Muhammad⁷

^{1,2,3,4,5}Department of CS & E, University of Engineering and Technology Lahore, Pakistan

saljoki78@gmail.com, DanishNazir34@gmail.com, tanzeelsarfraz1997@gmail.com, usamanasir517@gmail.com, maslam@uet.edu.pk

⁶Department of CS, CINVESTAV, Mexico City, Mexico

ammartin@cinvestav.mx

⁷Department of CS, Superior University, Lahore, Pakistan

mcs-s18-007@superior.edu.pk

Abstract— As the trend to shop online is increasing day by day and more people are interested in buying the products of their need from the online stores. This type of shopping does not take a lot of time of a customer. Customer goes to online store, search the item of his/her need and place the order. But, the thing by which people face difficulty in buying the products from online store is the bad quality of the product. Customer place the order only by looking at the rating and by reading the reviews related to the particular product. Such comments of other people are the source of satisfaction for the new product buyer. Here, it may be possible that the single negative review changes the angle of the customer not to buy that product. In this situation, it might possible that this one review is fake. So, in order to remove this type of fake reviews and provide the users with the original reviews and rating related to the products, we proposed a Fake Product Review Monitoring and Removal System (FaRMS) which is an Intelligent Interface and takes the Uniform Resource Locator (URL) related to products of Amazon, Flipkart and Daraz and analyzes the reviews, and provides the customer with the original rating. It is a unique quality of the proposed system that it works with the three e-commerce Websites and not only analyze the reviews in English but also the reviews written in Urdu and Roman Urdu. Previous work on fake reviews does not support feature to analyze the reviews written in languages like Urdu and Roman Urdu and cannot handle the reviews of multiple e-commerce Websites. The proposed work achieved the accuracy of 87% in detecting fake reviews of written in English by using intelligent learning techniques which is greater than the accuracy of the previous systems.

Keywords— Fake Reviews Detection, Text Classification, Natural Language Processing, Machine Learning, Bigrams, Term Frequency and Inverse Document Frequency, Urdu

I. INTRODUCTION

There are different ways to shop like you can buy a specific thing of your need by going to a store or mall. In this style of shopping the seller gives you the feedback of the product, you do not know whether he/she is giving a fake feedback or original. Because, it is upon seller honesty, how much the seller is true in his/her words and you have to carefully examine the product because you do not have any other option in examining the product [1]. If you don't pay attention in buying that product then it may be proved a waste for you. On the other hand, nowadays source of shopping has been changed. You can buy the products from the online stores of different brands. Here, you have to place the order without seeing and examining the original product. You read the reviews and buy the product. Therefore, you are dependent on the reviews about the product. These

reviews may be the original or fake. The customer wants to buy an original and reliable product, which is possible only when you get the original feedback related to that product.

Research shows that U.S. shoppers spend \$6 billion in Black Friday sale 2018 [2]. Americans spend 36% of the shopping budget online. In 2017, E-commerce stores earned \$2.3 trillion in sales and expected to reach \$4.5 trillion by 2021. Today, almost 12-24 million ecommerce stores are operating around the world. Study found that 61% of Amazon reviews that belongs to Electronics Category are fake [3]. There are some websites which are working to detect the fake reviews. Fakespot is an online Website that detects fake reviews using suspicious patterns and reviewers activity. As in the process of buying the product from the online stores you have to read all the reviews one by one to check for the quality of that product and to get a good quality product. It is a very time consuming process. Here this possibility also falls that the reviews may be fake or original.

However, if you are provided with the system in which you can find the original feedback and rating related to a specific product. Then, it is the source of satisfaction and reliability for you. In the proposed technique, the reviews related to a product for which the URL is given are extracted. After it, the system finds the fake reviews and finally by analyzing these reviews system finds the original reviews of the product. Previous researches detect fake reviews using different approaches including identification address, opinion mining and sentiment analysis, machine-learning approach. There are many researches available that detect the fake reviews related to English but no work is done so far that detect fake reviews for Urdu (highly spoken language in Asia) and Roman Urdu [4]. Therefore, we have proposed Fake Product Review Monitoring and Removal System (FaRMS) in which a customer can get the best possible item from the online store in a short time and with the original reviews associated with that product. This system gives you the original words of people related to the product with genuine reviews. Some popular products can get hundreds of reviews at some large merchant sites and FaRMS gives you the promising reviews by filtering fake reviews and then you can decide whether you want to buy or not.

The manuscript is organized as follows: Section II shows the literature review. Section III represents methodology of the proposed work. Section IV illustrates results of the proposed work and Section V illustrates the comparison of different classifiers. Finally, conclusions and future work is given in Section VI.

II. LITERATURE REVIEW

There is a lot of work that has been done on the fake reviews, fake news and fake social media IDs. Spotting groups of fake reviewers [5] finds the fake reviews and an individual fake reviewers group who are working on writing the fake reviews on the e-commerce Websites for promoting or demote the seller's product. Developers use the "Frequent Itemset Mining (FIM)" method to search the fake reviewers groups. The system uses the behavioral model and relational models to find the relationship among the fake reviewers groups which also called "spammer groups". Before this work, no labeled dataset is available so to check their method they produce a labeled dataset using expert human judges. This system uses novel relation based model called "GSRank" which finds the fake reviewers and relationship between the spammer groups. In this technique, the set of item I is the set of reviewer Id and each transaction is the set of Id who gives the review to a particular product and then the system uses the FIM method to find the groups who gives review of the multiple different products together.

The system of detecting fakes reviews of Yelp [6] analyzes that what Yelp is doing by filtering and analyzing these filtered reviews. There are two main approaches to filter reviews: supervised and unsupervised and in terms of features there are also two main types: behavioral features and linguistic features. Behavioral features perform well for their system rather than the linguistics features. Their starting point is the work of the Ott which uses the Amazon Mechanical Turk (AMT) to find the Turkers who are writing the fake reviews for the hotels. The system Fake Reviews Detection on Movie Reviews through Sentimental Analysis Using Supervised Learning Techniques is proposed with the aim to classify the movie reviews into positive and negative polarity by using the machine learning algorithm. The proposed system applies the sentimental analysis and text classification methods to detect the fake reviews related to the movies. The proposed work find out that Yelp may be using sentimental analysis but there is not hundred percent surety that they are using this technique to filter their reviews.

Fake reviews detection for the Yelp [7] is worked with the intention to filter the fake reviews from the original reviews as this is becoming the need of the hour. The proposed system classifier takes the reviews text and other information and produces the output whether the reviews are reliable or not. The data set which is used in this project is taken for the Yelp.com which is firstly used by the Rayana and Akoglu. They use 16282 reviews and split these into 0.7 training set, 0.2 dev set and 0.1 test set. Extracting predictive features from the reviews is the most challenging part of the project. Basically they extract two types of features: review-centric feature and reviewer-centric features. Firstly they count the percentages of each unigram and bigram tokens for fake and non-fake reviews. They then take out the top 100 unigrams and bigrams that have the most different percentages in fake and non-fake reviews. The second approach leads to the better performance because it processed all the unigrams and bigrams. They tested multiple algorithms of machine learning but by using the Neural Networks they achieved the highest accuracy of about 81.92%. This system is good in finding the fake reviews but still there is a need to improve the accuracy in filtering the reviews.

Spam reviews detection by using Temporal Pattern Discovery [8] is proposed to observe the reviews related to

the normal reviewers arrival pattern and fake reviewers arrival pattern and they observe that the normal reviewer arrival pattern is stable and uncorrelated to their rating pattern temporally. On the opposite side the spam attacks are usually bursty and either positively or negatively correlated with the rating pattern. The data set which they have taken is snapshot of a review Website on October 6, 2010. It includes 408469 reviews which are written by 343629 reviewers, which are written for 25034 stores of a Website. For each review they collect the following information like rating, postdate and whether it is an Spammer Review (SR) or not. In the evaluation process they select 53 stores each of which has more than 1000 reviews. Human evaluators make decision about the stores to be SR spam attack or not if two or more evaluators declared a store as SR spam attack than system considers the store a dishonest in its selling. Out of 53 stores 34 are suspicious one and the remaining are normal ones. Out of 34 store 22 stores have at least two votes for being suspicious. The recall related to the system is 75.86% which shows that the system detects most of the stores having SR spam attack. The precision related to the proposed approach is 61.11%. This proposed system is good in terms of the training of their model for finding the relation as the model is trained by using the large number of reviews contain in their dataset.

A technique to ranks the product [9] is worked to present a product ranking model that applies weights to product review factors to calculate a product ranking score. In this proposed system, the sentences that are not related with the quality of a product such as customer service or sentence related to the. In this paper the preprocessing is done by Support Vector Machine (SVM). First of all it removes the comments which are nor related with the quality of the product. Second stage describes the weights of the reviews based on the votes. The final stage calculates the overall ranking of the product. The ranking score is calculated by the relevance of the review with quality of the product, review content, and posting date of the review. They use 10-fold cross validation on the training set. In the evaluation process they use two measures to quantify effectiveness of the ranking model which are as following: correlation between the ranking method and the Amazon's rank and second is the Mean Average Precision (MAP), which is a very commonly used technique for evaluating ranking accuracy. As this system is finding the fake reviews by using the only two properties of the reviews but as per the future work describes in the paper more properties can be used to find out the fake reviews more accurately.

Fake review monitoring [10] system focuses on detecting spam and fake reviews by using sentimental analysis removes the reviews which have curse and vulgar words. In the proposed system web crawler is used to scrapped the data on the Website. In the preprocessing, the data is converted into the required format and then the fake reviews are removed from the mixture of original and spam reviews. Fake reviews are detected by the Fake Review Detector. Each review has to be passed from the classifier which calculates the review's sentiments score. Cosine similarity is used to measure this type of similarity. If the cosine value to be calculated is greater than 0.5 than the review is considered as the fake review. The developed system detected the 111 reviews to be faked out of 300 reviews. However, the data set which is used for training the model is very small which cannot find the suspicious patterns more accurately.

The previous works detect fake reviews by using IP address, opinion mining, sentiment analysis, and some of them uses machine learning approaches. In some approaches dataset is very small and other uses few properties related to the reviews to find the fake reviews. In the proposed system the large dataset of English reviews is used to train the model. In this way, the system can find the hidden patterns in the reviews more efficiently. The accuracy which is achieved by using the proposed technique is greater than the accuracy of the previous systems in terms of the English reviews. This system is also worked to detect the fake reviews of Urdu and Roman Urdu reviews.

III. FAKE PRODUCT REVIEW MONITORING AND REMOVAL (FaRMS) METHODOLOGY

Detecting fake reviews from product is very much important in this era. As there are two types of purchases in every e-commerce Website which are verified purchase and Non-verified purchase. Verified purchase means that the customer who was writing the review purchased the product from the online store and who did not receive the product at a great discount. To detect fake reviews, data is gathered from trip advisor, MTurk and Yelp. So that model can be trained in a best possible way.

For the verified purchase, it is an obvious thing that the customer has bought that product for which he/she is giving the review. Now, in this situation, if the customer leaves a positive review for the product and gives the rating of 1 or 2 than it is cleared that it is a fake review. That is why system uses sentimental analysis for the verified purchase. For the word like ‘good, excellent’ etc. the sentiment is to be positive and for the ‘bad, poor quality’ etc. type words the sentiment is to be negative.

For the Non-verified purchase anyone can give the review of the product without having any type of purchase history on e-commerce Website. Now in this situation, it is possible that the person who is giving the positive review related to the product also gives the good rating to that product. But, here this possibility also falls that the person may be giving the positive review and rating to increase the rating of that product and he is a spammer. In this type of reviews the technique of sentimental analysis is not worked properly because by using sentimental analysis it is considered as a genuine review. So that is the reason the proposed system is using another technique in which Support Vector Machine (SVM) is a classifier.

1. Approach for Verified Purchase

The proposed system takes the URL of e-commerce Websites like Amazon, Flipkart and Daraz products and based on that URL, scrape all the reviews related to that product. Stores all the reviews in a comma separated values (CSV) file. Load the CSV file and separate the verified reviews from non-verified reviews. Apply sentiment analysis on verified purchase reviews and add another column sentiment polarity. Sentiment is given based on polarity values given below.

Polarity > 0	(positive)
Polarity < 0	(negative)
Polarity = 0	(neutral)

Assigning the positive, negative or neutral to each review has been done at this stage. After it, the system checks for the rating and sentiment of the review. If the rating of the product is 4 or 5 and the sentiment of the review is negative or neutral then the review is considered to be fake. If the sentiment is positive then the review is considered to be genuine one. But, if the sentiment is positive or neutral then the review is considered to be a fake one. If the review has the rating of 3, and the sentiment of the review is either positive, negative or neutral it is consider to be a genuine one.

2. Approach for Non-verified Purchase

The technique to detect the fake reviews of the non-verified purchase, we use another approach. For this first step is to collect the text related to the reviews of the product. After gathering textual data, the system applies preprocessing on each review related to the product and then extracts features from the reviews. After features extraction, the next step is to apply the SVM model and after it the results corresponding to the genuine or fake reviews is to be shown to the user. Fig.1. shows the complete architecture of FaRMS.

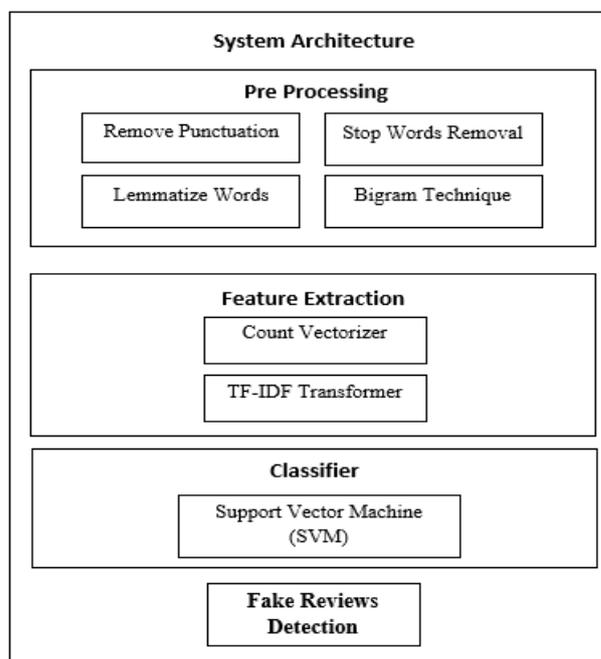


Fig.1. FaRMS Architecture

2.1. Collection of Dataset

The dataset of English reviews that the system is using to train the model was the Deceptive Opinion Spam Corpus¹, which consists of both the deceptive and genuine reviews about hotel collected from Trip Advisor and MTurk. Data consists of 1600 hotel reviews of which 800 reviews are collected from Trip Advisor and are genuine whereas other 800 reviews are collected from MTurk that are deceptive. Out of 800 reviews 400 reviews have the positive polarity and 400 have negative polarity. The accuracy of the proposed system by using this dataset is 90% but this is a very small dataset to train the model in good sense.

¹ <http://myleott.com/op-spam.html>

The dataset that is discussed before is quite small. Therefore, system demands the English reviews related dataset that contains a large number of reviews. For that, the dataset is collected from the Yelp² that is labelled (review is As, Yelp has a filter that automatically detects fake reviews. The proposed system is trained by using this dataset [11] in order to train the model to detect fake reviews. The dataset contains two files metadata and content data. The metadata contains 359,052 rows and content data contains 358,957 rows. The dataset contains 36,860 deceptive reviews and remaining as genuine reviews.

For Urdu and Roman Urdu reviews dataset, English reviews of Yelp are translated manually into Urdu and then translated from Urdu to Roman Urdu. As, Google translator cannot translate English to Urdu language with much accuracy. Therefore, each review is translated manually keeping in mind the semantics of each review. This might have an impact on the semantics of the reviews and as a result, accuracy can vary. Corpus of total 400 reviews of both Urdu and Roman Urdu is made manually. Out of 200 of the reviews are deceptive and 200 are truthful reviews.

2.2. Cleaning of Data

The Deceptive Opinion spam dataset contains no cleaning of data because it is in the presentable form. The Yelp dataset requires data cleaning because the rows in metadata and content file are not equal. The proposed technique clean the data by removing rows from metadata that are not in content file by matching the date, user id and product Id from both files. In this way, system has the same number of rows in both the files and then the data is combined. Now, the system is ready for further processing.

2.3. Loading the Data

Both of the datasets are in CSV format so Pandas library of python is used to load it into the desired python Integrated Development Environment (IDE). Yelp dataset contains unequal distribution of deceptive and truthful reviews so the system is loaded with only the 110,580 rows data in which 36,860 reviews are deceptive and remaining are genuine reviews. To balance the dataset system framework duplicate the deceptive reviews so the system contains the equal 147,440 reviews half of them are deceptive and half are genuine reviews. In this way, overfitting is avoided and model is trained accurately. Urdu and Roman Urdu dataset are loaded in the same way.

2.4. Visualization of Data

After visualizing, the data system find out that the length of fake reviews are long and contains words that are more positive, more punctuation, and repetition of words.

2.5. Splitting the data

For splitting the data into training and testing one, proposed technique uses the most common form of splitting that is 20% for testing and 80% for training.

2.6. Data Pre-Processing

Before representing, the data using the n-gram model and add features to it, first the system need to do some

refinements to the present data that includes removing punctuation and stop words, convert all the data into lower case. This helps to focus only on the actual data that gives more information rather than the information that only adds noise in the model. The system uses the functions to remove all the punctuation, stop words, and then convert all of the remaining words in lower case. After that, the propose technique applies lemmatization and then uses the bigrams technique.

2.6.1. Punctuation Removal

The first step in preprocessing is to remove the punctuation from the text. Punctuations are the marks such as full stop, comma, semi-colon, hyphen used to separate sentences from one another to clarify the actual meaning. These marks were removed from each review.

2.6.2. Stop Words Removal

Stop words are the words that are used a lot in the sentences to connect them. Stop words only create noise in the feature extraction so these words should be removed before doing text classification. Articles, Preposition and some pronoun are considered stop words. System is removing These stop words such as is, are, am, he, from, who etc. Urdu and Roman Urdu, stop words are also removed for better training of data and better results. These words are removed from each review and then the review is passed on to the next step.

2.6.3. Lemmatization

Lemmatization is the basic text classification method for English text. The goal is to convert the word into the common base form. It is the grouping of different inflected form of words so they can be used as a single word. It is basically to determine the lemma of the given word. For example, the verb 'to walk' may appear as walk so walk is basically the lemma. The system technique is using lemmatization to make the classifier faster and efficient. The word is reduced to its lemma and then the system is moved to the next step in preprocessing.

2.6.4. N-gram Modeling

N-gram modeling is the most popular feature identification approach in natural language processing. N can be any number. The most important n-gram modeling in text classification is the word based or the character based. System uses this model in order to generate features to classify the reviews. In the proposed approach, bigrams (N = 2) is using to find whether the reviews are deceptive or truthful. After converting each review in bigram form, system is done with the preprocessing step and now system is ready move to the feature extraction technique.

2.7. Feature Extraction

Feature Extraction is the most difficult task that is the main reason behind your classifier accuracy. Irrelevant data can reduce the performance and accuracy of the classifier. Thus, it is better to eliminate the irrelevant data before extracting the features. System is studied by two feature selection methods: Term Frequency (TF) and Term Frequency-Inverse Document Frequency (TF-IDF) for the selection of

features related to our dataset. In the proposed approach, system uses Count Vectorizer that converts each review into bag of words and is used to tokenize the set of words described in the reviews and after it system applies TF-IDF transformer. The model converts the collection of text documents into a matrix of token counts. It is 2-dimensional matrix where 1-dimension represents the vocabulary and other dimension of actual document as described in Table 1.

Table 1. Count Vectorizer Overview

	Word 1	Word 2	...	Word N
Review 1	0	2	...	1
Review 2	0	1	...	1
...	1	0	...	2
Review N	2	1	...	0

Since, there are a lot of zero involved in this matrix so it is called sparse matrix (have many zero values).

The TF-IDF Transformer is the weighted metric used in text mining and is used to measure how important is the word in that dataset. Importance of the word increases based on how many times the word appears in the dataset. Each word is assigned a respective TF-IDF score. For a word t in document d , the weight $W(d, t)$ of the word t in the document d is given as describe in Eq. 1:

$$W(d, t) = TF(t, d) \times \log\left(\frac{N}{DF(t)}\right) \dots\dots\dots(1)$$

Where, $TF(t, d)$ is the number of occurrence of word t in document N is the total number of documents (reviews) in the dataset $DF(t)$ is the number of documents (reviews) containing the word t .

2.8. Classification Process

Fig. 2 shows the classification process of FaRMS that how the proposed system is working to classify the reviews into genuine and fake ones. It starts with collecting the data, the next step is to preprocess the data including removing punctuation and stop words from the text of the reviews, then system converts the text into lower case, lemmatize the word to its lemma and the last step of the preprocessing the system uses the bigram technique to convert the text into bigrams. After preprocessing, extract features using Count Vectorizer that converts each review into 2-D matrix and then apply the TF-IDF transformer that gives weight to each word. After the feature selection, the last step in the classification process is to train the classifier. The proposed architecture is tested by applying three different supervised machine learning algorithms including SVM, Naïve Bayes, Logistic Regression but SVM outclass the remaining classifier with its performance and compete the other algorithms in terms of the results.

2.9. Predictions and Evaluation

After successfully training the data, system is applied with the testing data to predict unseen data in order to find out whether it is deceptive or genuine. On the Deceptive Opinion Spam dataset of hotel reviews, system has achieved the accuracy of 90%. In Yelp reviews dataset proposed technique obtain 87% accuracy with bigrams feature. In the dataset related to Urdu reviews, system has achieved the maximum accuracy of 70% by using the SVM classifier with bigrams feature. In Roman Urdu dataset, system has

achieved the maximum accuracy of 69% by using the same SVM model and bigram technique. Classification process of reviews into truth and deceptive ones is show in Fig.2.

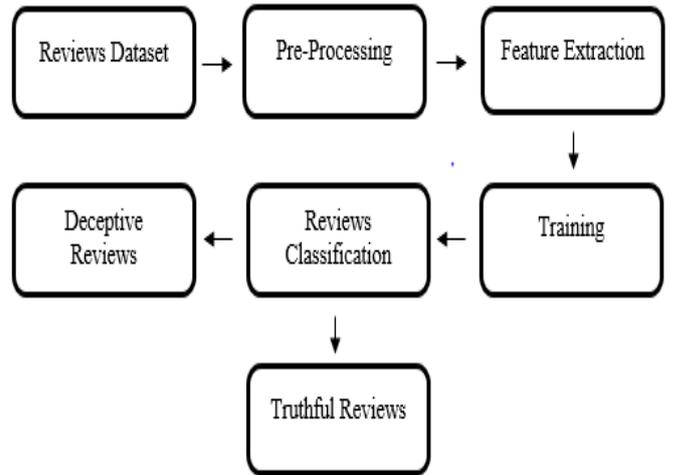


Fig. 2. FaRMS Classification Process

IV. RESULTS AND VISUALIZATION

Yelp Dataset: The proposed architecture uses the SVM classifier to train the model that consists of 147,440 rows. Confusion matrix of this dataset is given in Table 2.

Table 2. Confusion Matrix of Yelp Reviews Dataset

	Predicted False	Predicted True
Actual False	13,542	1185
Actual True	2700	12,061

Urdu Dataset: System uses the SVM classifier to train the model that consists of 400 rows. Confusion matrix of this dataset is shown in Table 3.

Table 3. Confusion Matrix of Urdu Reviews Dataset

	Predicted False	Predicted True
Actual False	25	8
Actual True	16	31

Roman Urdu Dataset: The proposed architecture uses SVM classifier to train the model that consists of 400 rows. Confusion matrix of this dataset is shown in Table 4.

Table 4. Confusion Matrix of Roman Urdu Reviews Dataset

	Predicted False	Predicted True
Actual False	29	15
Actual True	10	26

V. COMPARISON BETWEEN DIFFERENT CLASSIFIERS

The proposed system is tested with three different classifiers to find the best one and it is observed that SVM performs best as compared to the others three classifiers.

Yelp Reviews. Yelp reviews [11] dataset consists of 147,440 reviews. 80% data is used for training and remaining 20% for testing. Performance on different classifiers is shown in Table 5.

Table 5. Comparison of Classifiers for Yelp Reviews

Classifier	Accuracy
Support Vector Machine (SVM)	87%
Naïve Bayes	85%
Logistic Regression	81%

Urdu Reviews. The dataset of Urdu reviews consists of 400 reviews. 80% of the available data is used for training and the remaining 20% data is used for the testing. Performance on different classifiers is shown in Table 6.

Table 6. Comparison of Classifiers for Urdu Reviews

Classifier	Accuracy
Support Vector Machine (SVM)	70%
Naïve Bayes	68%
Logistic Regression	68%

Roman Urdu Reviews. The dataset related to the Roman Urdu reviews dataset consists of 400 reviews. 80% data is used for training and remaining 20% for testing. Performance on different classifiers is shown in Table 7.

Table 7. Comparison of Classifiers for Roman Urdu Reviews

Classifier	Accuracy
Support Vector Machine (SVM)	69%
Naïve Bayes	66%
Logistic Regression	65%

VI. CONCLUSION AND FUTURE WORK

In the proposed work, dataset is developed that contains Urdu and Roman Urdu reviews. It is difficult to detect fake reviews by yourself. So, n-gram approach is used to detect fake reviews for multiple languages. It is observed that the text categorization with SVM classifier is best approach for the detection of fake reviews. Now a days, as the technology is growing day by day and there are so many Websites and applications available in the online market by which seller can sell their products and on that products there are millions of reviews available. There are some organizations posting fake reviews for the products of the seller in order to increase or decrease the rating of the products. Therefore, the system is proposed that detects the fake reviews in multiple languages including English, Urdu, and Roman Urdu, classify the reviews in genuine. It helps

the user to get the products from Daraz, Flipkart and Amazon with the satisfaction of their mind and pay for the good quality product.

As, there are a lot of e-commerce stores like AliExpress and Alibaba which have reviews of multiple languages. It would be great if the proposed system finds the way, to process and filter the reviews for other multiple languages.

Funding. This research is funded by the National Research Program for Universities (NRPU), Higher Education Commission (HEC), Islamabad, Pakistan, grant number "20-9649/Punjab/NRPU/R&D/HEC/2017-18".

REFERENCES

- [1] A. Sinha, N. Arora, S. Singh, M. Cheema, and A. Nazir, "Fake Product Review Monitoring Using Opinion Mining," vol. 119, no. 12, pp. 13203–13209, 2018.
- [2] Torbet, Georgina. "U.S. Customers Spent over \$6 Billion on Black Friday Purchases." Digital Trends, Digital Trends, 25 Nov. 2018, www.digitaltrends.com/web/shopping-totals-black-friday/.
- [3] Sterling, Greg. "Study Finds 61 Percent of Electronics Reviews on Amazon Are 'Fake'." Marketing Land, 19 Dec. 2018, marketingland.com/study-finds-61-percent-of-electronics-reviews-on-amazon-are-fake 254055.
- [4] K. Khan, W. Khan, A. Rehman, A. Khan, Asfandyar. Khan, A. Ullah Khan, B. Saqia, "Urdu Sentiment Analysis," (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 9, No. 9, 2018.
- [5] A. Mukherjee, B. Liu, and N. Glance, "Spotting Fake Reviewer Groups in Consumer Reviews," 2012.
- [6] A. Mukherjee, V. Venkataraman, B. Liu, and N. Glance, "What Yelp Fake Review Filter Might Be Doing?," Aaai, pp. 409–418, 2013.
- [7] Z. Wang, Y. Zhang, and T. Qian, "Fake Review Detection on Yelp Dataset and features," pp. 1–6.
- [8] S. Xie, G. Wang, S. Lin, and P. S. Yu, "Review spam detection via temporal pattern discovery," p. 823, 2012.
- [9] C. Paper, "Mining millions of reviews : A technique to rank products based on importance of reviews Mining Millions of Reviews : A Technique to Rank Products Based on Importance of Reviews," no. November, 2015.
- [10] V. K. Madhura N Hegde, Sanjeetha K Shetty, Sheikh Mohammed Anas, "Fake product review monitoring," Int. Res. J. Eng. Technol., vol. 05, no. 06, p. 4, 2018.
- [11] S. Rayana and L. Akoglu, "Collective Opinion Spam Detection : Bridging Review Networks and Metadata," Sigkdd, pp. 985–994, 2015.