ISSN: 0369-8963

Spam Review Identification Metrics for Distinguishing the Honest Reviews from Fake Reviews

DR.GOBI NATESAN,
Assisstant Professor, Department of CSE,
Dr. Mahalingam College of Engineering and Technology,
Pollachi, Coimbatore, Tamilnadu, India-642003.

MRS.K.BALASARANYA,
Assistant Professor,
Department of Computer Science and Engineering,
R.M.D. Engineering College,

ABSTRACT:

Sentiment Analysis (SA) systems are quite popular nowadays since most people trust them to make decisions on products, services, and news analytics, among other things, based on the reviews expressed by the end users. Customers will employ Sentiment Analysis tools to select new products from a variety of options available. Manufacturers can use the implemented system to known about their products' strengths and weaknesses.

At present the sad aspect is that, many spammers post the irrelevant or fake reviews about certain products to increase or decrease its market share among others. Sentiment Analysis systems have a difficult time deploying methodology to identify whether each review is either honest or spam and also to known whether it was made by individual spammers or spammer groups after experiencing the products.

The suggested system will provide solution to the problem faced above by utilizing Text pre-processing as the best place to start when it comes to increasing the overall effectiveness of sentiment analysis systems. Subsequently, an innovative spam review detection approach namely Spam Review Identification Metrics (SRIM) is implemented based on several factors determined through review level and reviewer level characteristics to classify the review as honest or fake present in the Review dataset. Multilayer Perceptron (MLP) classifier is used to identify the given review as spam or honest and performs well when compared to other classifiers like Naïve Bayes and Decision Tree techniques. An interesting observation is that, although the performance of positive sentiment identification by the Naïve Bayes and Decision Tree outperforms well by 2.28% and 1.45% more than MLP. However, MLP produced good results for negative sentiment related reviews by 2.36% and 4.37 % more than NB and DT methods.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

INTRODUCTION:

Due to increased growth of E commerce, purchases and selling of the products through online is popularly acknowledged. Some statistics says that almost 40% of the internet users worldwide buy their household products online through E-Commerce application. The advancement of internet makes the individual users to express their views and feelings about their purchased product on the web through written reviews. A study conducted during the year 2014, reveals that online reviews are very significant for buyers, as over 90% of customers used online reviews to help them make decisions before making the payments to any merchant. A strong positive and negative opinion about the products will significantly influence customers towards accepting and rejecting the products respectively. It is very common that, positive opinion for any service or product can attract large customers which lead to substantial financial gains. Similarly, the negative opinion may lead to downside of the product purchase history. There is a possibility that certain corporates appoint a group of spammers for writing the fake reviews or spam reviews about certain products which may lead to the fluctuation in the product market value. It's difficult to detect and eliminate fake reviews from the data set to keep spammers from being subjective to their unethical behaviour. The available of spam reviews in the dataset and spam reviewers used for writing the spam reviews has been a hot issue of controversy in recent years, and various researchers have dedicated significant time and effort to analysing, detecting, and eliminating fraudulent reviews from the Dataset.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

WORK ON DETERMINATION OF FAKE REVIEW PRESENT IN THE REVIEW DATASET

Due to lot of improvement in sentiment analysis during last decade, it is easy to determine the polarity of the review based on numerous approaches. But recently there is a trend introduced by the intruders to write the fake reviews about the product in order to increase or decrease the market value of certain products. It is difficult for the researcher to determine whether the dataset includes any fake reviews through the effective SA system. Lot of algorithms and methods are introduced by the researches to identify and remove the fake reviews from the dataset.

Many organizations and enterprises use SA systems to determine the positive and negative aspects mentioned by the customer for the products based on reviews for improving their sales through the intelligent algorithms developed by various researchers. However, all reviews given by the reviewer are not true reviews as some spammers ultimately aim to promote or demote the product value as discussed by Dinesh Kumar Vishwakarma *et al.*(2019). Many researchers focus on detecting the fake reviews and build the trustworthiness system for the organization for producing an effective review analysing system. The various Text Mining and Opinion Mining techniques are implemented to measure the honest value of a review discussed in Chaowei Zhanga *et al.*(2019). Manali S.Patil *et al.*(2012) proposed a novel approach where untrustworthy reviews are identified using n-gram model and the Spam review detection is based on features selection techniques. The above model identifies the spam review written by using same credential for different products simultaneously and different credentials for writing spam reviews for the same products. The accuracy of predicting the spam review is 65% comparing to its previous model.

Eka Dyar Wahyuni *et al.* (2016) proposed an Interactive Computation Framework (ICF) which is used to detect fake reviews from the dataset based on review content and rating attribute of a review. This method is used to measure the honest value and trustworthiness of the reviews written by the reviewer and maintain

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

the reliability of the products in the market. Proposed method has a better accuracy compared to the result from ICF. This method has a drawback of predicting the review by measuring the time taken to predict the spam score. The accuracy of the system is 72% as against the ICF accuracy of 66%. Julien Fontanarava *et al.* (2017) suggested the prediction of spam that illustrates the impact of distinct features through cumulative distribution functions approach and it is possible to identify the more relevant features with respects to fake reviews as with the most different cumulative distribution functions between the classes. It also considers the different ways the spammers used to write the reviews for the same product during a particular time interval (Burst) to detect the singleton fake reviews. The accuracy of predicting the fake reviews through this approach is 80.6%. Doaa Mohey El-Dia Mohamed Hussein (2018) apply the sentiment classification algorithm using SVM with stop word removal method to detect the fake reviews present in the dataset having the accuracy is 81.2%. Table 1 represents the comparative analysis of various Fake review Detection approaches. It clearly indicates that certain authors focus on Review information and Reviewer activity to classify the review as spam or honest. Few authors considered only the Review information for classifying the spamicity of the reviews which is considered as less significant when considering the authors significance. Certain authors considered burst pattern to predict the class of the reviews as spam or honest along with the rating behaviour which produces good accuracy over the other techniques discussed by different authors.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

Table 1 Comparison among different approaches towards Fake review Detection

	Type of *Information used for Analysis*		Approaches Used						Graph Modeling
	Review Content	**Reviewer Activity**	**Identification of Similar reviews**	**Categorization of Text analysis**	**Detection**	**Reviewer Identification**	**Burst Pattern Discovery**	**Review rating**	
Ott *et al.* 2011	X	-	-	Word n-grams	X	X	-	-	-
Wang *et al.* 2011	X	X	-	-	-	-	-	-	X
Fayazbakhsh andSinha 2012	X	X	-	-	-	-	-	-	X
Xie *et al.* 2012	X	-	-	-	-	-	X	-	-
Ott *et al.* 2011	X	-	-	Word n-grams	-	-	-	-	-
Mukherjee *et al.*2013	X	X	X	-	-	-	-	X	-
Akoglu *et al.* 2013	X	X	-	-	-	-	-	-	X
Fei *et al.* 2013	X	X	X	-	-	-	X	X	-
Lin *et al.* 2014	X	X	X	-	-	-	-	-	-
Banerjee and Chua 2014	X	-	-	-	-	X	-	-	-
Hernandez-Fusilier *et al.*	X	-	-	Char n-grams	-	-	-	-	-
Savage *al.* 2015	X	X	-	-	-	-	-	X	-
Heydari *et al.* 2016	X	X	X	-	-	-	X	X	-

https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

From the above Table 1, it is concluded that spam review is detected at review level and reviewer level along with burst pattern and rating behavior. So, the proposed Spam Review Identification metrics system considered all the above parameters along with some additional features to produce good accuracy than all the above literature works.

PROBLEM DEFINITION

Opinion spamming is the process of using illegal methods to produce bulk amount of fake reviews about a specific services, utilities or product in order to promote or degrade the product's quality. Fake reviews or sometimes termed as spam reviews are created for the above purpose, and the author who publishes the fake review is called as spammer. The following aspects are the great impact of spam reviews and spammers:

- Negative spam reviews against any product will make the business result in less profit due to underserved damage reputation by its competitors
- Customers may be misled about the quality of the product they have purchased.
- An increase in the number of spam reviews may frustrate internet users.

The researchers used a variety of indicators to detect spam in the review dataset successfully, which includes combination of variety of spam identification metrics and techniques to categorize the honest review from fake review. To identify fake from genuine reviews, a score is derived from several metrics using quantitative measurements such as review author, and a review history score is assigned to each review under examination.The Proposed work, namely Spam Review Identification Metrics (SRIM) are used to identify the spam reviews available in the Dataset for Spam Review is based on,

- Implementation of machine learning classification algorithm to classify the review textual contents.
- Detecting the Fake reviews by combining the reviews and reviewers based techniques.
- Determination of the suspiciousness behaviour and reputation of Author through their history reviewing process.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

BASIC APPROACHES TOWARDS THE DETECTION OF SPAM REVIEW

The following are three major techniques used for identifying spam reviews in review data:

- Detection of Spam review.
- Identification of Spam reviewer.
- Spammer groups detection.

Techniques for Spam Reviews Extraction via Review Content Similarity

Large numbers of supervised learning methods are employed to detect the fake reviews and they are based on supervised training data set. Review Similarity is the measure to identify the relevance between the two review sentences of similar or different products.

Techniques for detecting the Spammer and Fake reviewer based on Proliferation

Detecting the review spammers is primarily examined through their behavioural patterns. Spammers and non-spammers display differential behavioural patterns which lead to the creation of two clusters of reviewers namely spam and non-spam reviewers. The examined characteristics and behavioural aspects tend to calculate author related spamicity score having the values of 0 or 1 where Score 0 represents the non-spamming activity and Score 1 indicate the spammers and identify their review types through Content Similarity, Maximum number of reviews and Review burstiness.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

Techniques used for Spammer Groups Detection

With case studies for detecting fake reviews and opinion spammers, a significant number of researchers have devoted their time to detecting spammer groups. Spammer groups are made up of reviewers who collaborate to promote or degrade a group of target products by writing bogus or fake reviews. Supervised approaches are derived to detect the spammer group based upon group features like Time Window, Content Similarity and Content Size. Recently, a Group Spam Rank method (Mukherjee *et al.* (2013)) compiles all the aspects of spammer groups based on the probability of spamming. GS Rank outperforms supervised classification and regression to detect the spammer groups.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

TECHNIQUES FOR THE DETECTION OF SPAMICITY THROUGH THE COMBINATION OF REVIEW AND REVIEWER CHARACTERISTICS

In order to identify spam activity in reviews, the proposed technique aims to include strategies that make use of all accessible information regarding reviews and reviewers. Each review is given a rating that determines the level of spam, making SRIM easier. When the spam rating of a review exceeds a specific threshold, it is considered spam. For example, a review of a recently released product with a rating of more than 4.5 on a 5-point scale may be classified spam. Each discrete approach assigns a score to the reviews, and the final spamicity score to calculate SRIM is derived by aggregating all of the different metrics' values for each review in the dataset.

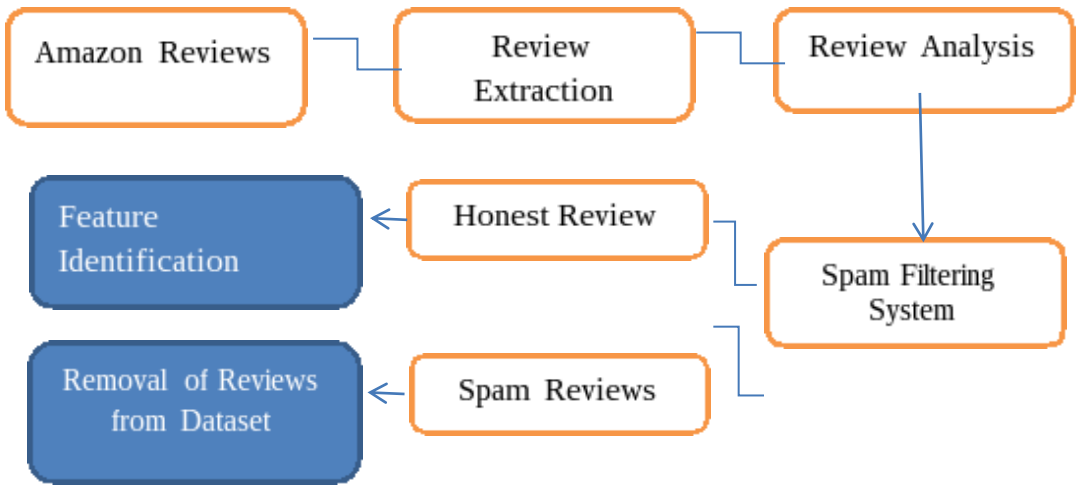

Figure.1 Generic process of detecting the Honest and spam Reviews present in the Dataset

Figure 1 depicted the process of determining if a review was spam or not. The suggested approach's initial step is to extract and collect all reviews for the entity (such as Amazon Product or Hospital Services) using pre-defined information such as content, rating,

https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

creation date and time, and author. The review analysis process then begins with a variety of approaches to comprehend the two levels of the problem, such as the review and the reviewer. A spammer is likely to produce false content about any product or services and it will therefore affect the behaviour of the decision making process and also about the authenticity of the review. After the review analysis phase, the next step is to determine the spam content present in the review through spam filtering system by analysing the behaviour of the review and reviewer.

Finally, SRIM is measured through Spamicity score by analyzing both review (*r)* and reviewer (*a)* indicator using the following scoring function in Equation (Eq 1).

$$\text{Spamicity}(r) = \text{Review Features}(r) + \text{Author Behavior}(a) + \text{Reputation}(a) \quad \text{(Eq 1)}$$

As illustrated in Figure 2, a review is labeled spam if its spamicity score (calculated using SRIM) exceeds a specific threshold, but a review that clears the threshold value is treated as spam review. The forthcoming sections that follow outline the methods and reviewing characteristics that go into calculating the spam score for a review. A threshold value of 25 is set for the SRIM system rather than other values set by the previous works, since the proposed work includes three important parameters namely Content Length, Review Relevancy Rate and Number of Review for effective identification of spam reviews. In general, the range of values for most of the parameters are between 0 and 2 but only three parameter [RRR, NR, AvgP] values are highly variable across multiple ranges based on the reviews given by the genuine user or spammers. So, in order to determine the values of the spamicity of a review, it is fixed to have a threshold value of 25 for the proposed system. Suppose, if the calculated spam score exceeds a threshold value of 25, the class of reviews is considered as spam else the review is treated as an honest one as indicated through the Figure 2.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

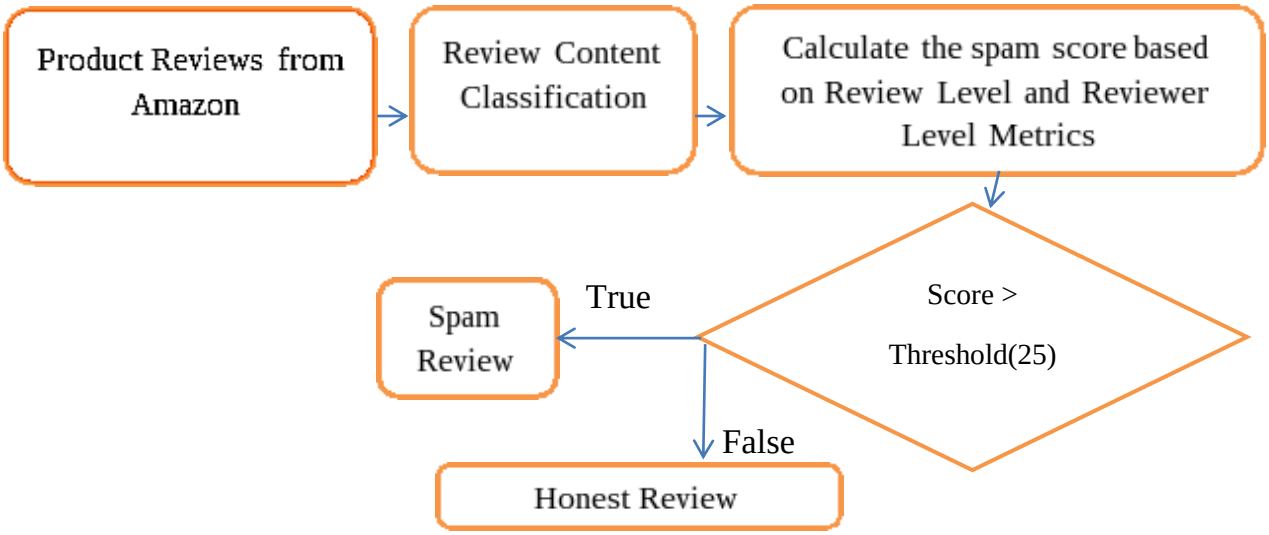

Figure 2 Flow diagram for Proposed Fake Review Detection using Spam Review Identification Metrics

Review Content Classification

The first approach to be investigated at is identifying the review content as spam or genuine. Machine learning algorithms have previously demonstrated promising results in a variety of text classification scenarios, including spam review identification. The review classification approach used in the suggested detection system is based on the Multilayer Perceptron (MLP), a machine learning algorithm, which helped significantly to the detection of review spam.

MLP is a feed forward artificial neural network model that translates input data to appropriate outputs (Figure 3).The most essential activation function in MLP is sigmoid, which is given in Equation (2).

$$y(q_i) = (1 + e^{-q})^{-1} \quad \text{(Eq 2)}$$

Where $y(v_i)$ is function in which the 'q' indicates the weighted aggregate of the input synapsis and the ‘i' represents the each individual neuron output.

https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

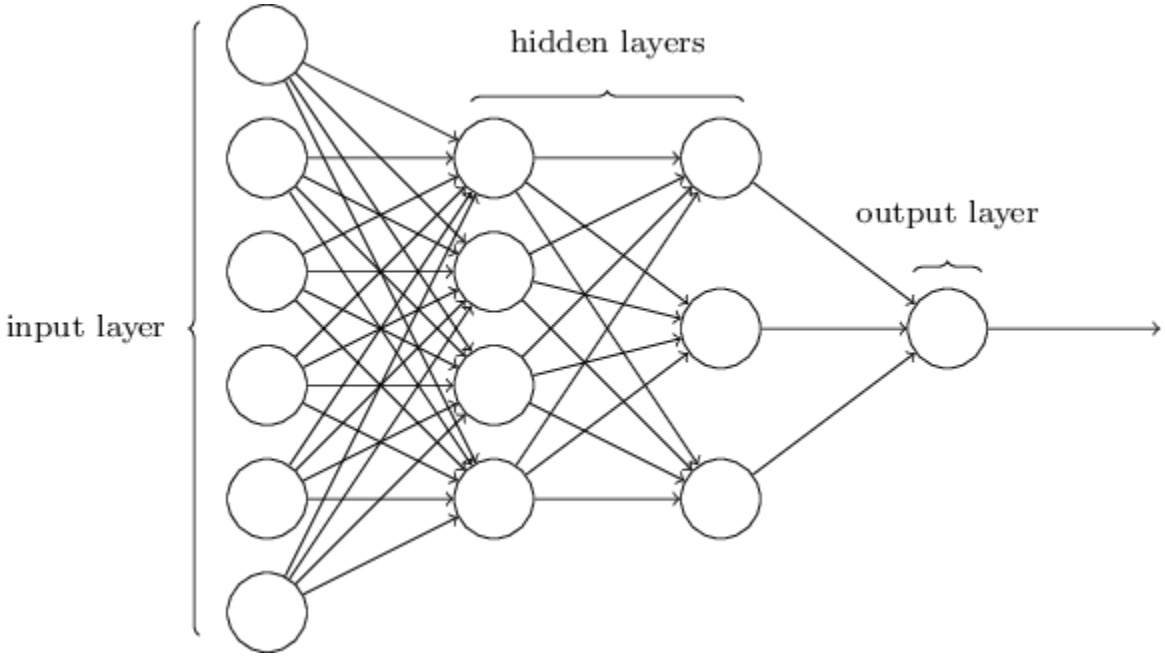

Figure.3 MLP Network Structure with Multiple hidden layer

Back-propagation, a supervised learning approach, is used to train MLPs. Back propagation is a popular method for training artificial neural networks, and it's frequently used in conjunction with an optimization method like gradient descent. The technique repeats a biphasic cycle of propagation and weight updating at each step. When an input vector is received by a network, it is propagated layer by layer until it reaches the output layer. A loss function is used to compare the network's output to the expected output, and an error value for each of the output layer's neurons is calculated. The error values are then reversibly transferred, beginning with the output and ending with an error value that roughly equals each neuron's contribution to the original output. Using these error numbers, back propagation determines the gradient of the loss function in proportion to the weights in the network. In the second phase, this gradient is passed to the optimization technique, which uses it to update the weights and minimise the Loss function.

During back propagation, nodes weight $\varepsilon(n)$ (Equation (3)) have to be modified in accordance to minimize the overall error in the output:

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

$$(n) = \sum_j e^2(n) \quad \text{(Eq. 3)}$$

$(n) = (n) - yj(n)$ represents the j-th node's error for the n^{th} training instance, where 't' representing the target value and 'p' representing the perceptron's generated value. Equation (4.4) is used to indicate the change in each weight.

$$\Delta wji(n) = -\eta\ (\partial s(n)/\partial vj(n))*yi(n) \qquad \text{(Eq. 4)}$$

Where Δw_{ji} represents the weighted average of weight from neuron *j* to *i*. Here $y_i(n)$ is the previous neuron's output and η be the learning rate is a number that permits the weights to change into a suitably fast response without causing an oscillation and listing the previous neuron's output. The calculation is performed based upon the gradient descent, a first-order recursive optimization process that finds the local minimum of the function in question by taking steps proportionate to the function's gradient at the current position. The above computation is based on the local area that is induced and simplified using Equations (5) and (6) .

$$\frac{-\partial\varepsilon(n)}{-\partial v_i(n)} = e_j(n)\varphi'\left(v_j(n)\right) \qquad (Eq.5)$$

$$\frac{-\partial\varepsilon(n)}{-\partial v_i(n)} = \varphi'\left(v_j(n)\right)\sum\nolimits_k -\frac{-\partial\varepsilon(n)}{-\partial v_k(n)} w_{kj}(n) \qquad (Eq.6)$$

Where φ' represents the non-variable derivative of the activation function.

When the network is trained, the middle layers' neurons organise themselves so that different neurons learn to detect different features of the whole recording region. After training, when an arbitrary input pattern with noise or gaps arrives, neurons in the network's hidden layer respond with an active output if the incoming input contains a pattern that the individual neurons have learned to recognise during exercise.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

Each evaluation of a certain product is transmitted to a trained MLP classifier, and label is attached to it. The classification will results in binary output that is Fake or genuine review and it is based on the categorization process of the MLP classifier. When the substance of a review is examined, a score is assigned, with 1 indicating spam and 0 indicating a genuine review. Equation (7) can be used to express the Classification Factor (CF)).

$$CF(r) = \begin{cases} 1, & label = spam \\ 0, & label = honest \end{cases} \qquad (Eq.7)$$

The method outlined seeks to make use of an important aspect of evaluations, namely their content. It is significantly reliant on the training data while conducting serious content analysis research. However, the writing styles of spammers and genuine users can differ, thereby limiting the usefulness of Review ratings.

Review Relevancy Rate

There are reviews in the dataset that are unrelated to the product, expressed through advertisement or a link. The system used to identify the relevance between the review content and the product under investigation is referred to as the Review Relevance Rate. To make customer purchases easier, each product has its own subject or content to understand about its features, such as product specification, functionalities, and its usage. Equation 8 is used to compute the Review Relevancy Rate (RRR).

$$RRR(r) = e^{((W(s) \cap W(r)/W(s))} - 1 \qquad (Eq.8)$$

W(s) contains the keyword list related to product, while W(r) contains keywords related to a review. A review is considered as spam if it has a high relevance rate.

https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

Content-Length

The length of a review's content can also be used to identify fake reviews. When a review is too short, it implies that the reviewer did not think the product's experience was important. As a result, these types of reviews are worthless for analysing review data. The review's Content's Length (CL) calculated by using the Equation (9).

$$CL(r) = \begin{cases} 1, r.length \leq \lambda \\ 0, r.length > \lambda \end{cases} \qquad (Eq. 9)$$

where *r. len* denotes review's *r length*, and t indicate the threshold value to predit the accuracy of reviews' content length.From the past research, it is concluded that a review has less than 5 words long, the review should flag it as spam because the author is less serious about the product.

Rating Deviation

Rating Deviation (RD) plays a significant part for calculating the spamicity score. It identifies how one review deviates from another through its activity analysis by the same reviewer. Spammers aim to promote or demote certain products and brands through varying their rating values.

Before assigning a value to a service or product by an Reviewer , the presented technique considers a Review r and analyses the deviation of the review from the average score for the specified product. The deviation d from the mean of a reviewing score s is calculated using the equation below.

$$d = |s - S_{mean}| \qquad (Eq. 10)$$

To avoid analyzing differences from an average score that is some factors have already affected rating given by the same author, which could damage the effectiveness of the system. However, the overall average score does not include all reviews written by the same author. Finally, using

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

Equation (11) for a specific product, Rating Deviation point RD (r) of a rating r is derived by normalizing according to the grading scale Nscale.

$$RD(r) = \frac{d}{N_{scale}} \qquad (Eq. 11)$$

Measuring the reviewer activities through number of reviews written and their activeness

As previously stated, a review's total spam score is determined by the reviewer's behavior as well as the reviewer's history being experimented with. The quantity of reviews a reviewer leaves for a single product is one of the most clear signs of whether they are genuine or malicious. Spammers attempt to mislead the product's feedback section by writing a high number of false reviews in an attempt to drown out honest feedback. The number of reviews NR (a) that a reviewer writes for a product also factors into the calculation of the Spamicity Score.

A fake reviewer have a tendency to write a large number of reviews in quick time may contribute towards the creation of Products' peak reviewing activity. An author's activity in bursty time slots, that is, the total amount of "bursty" reviews posted by the author helps to calculate the review spamicity score using the below formulae:

$$BuA(a) = \begin{cases} 1, Bursty\ Reviews > 2 \\ 0, Bursty\ Reviews \leq 2 \end{cases} \qquad (Eq. 12)$$

The Equation (12) states that if an author post more than two reviews for a particular product, there may be possibility of considering those review as spam.

The following sections talks about the author related features which include the authors previous activity during review writing process.

Reviewer Burstiness

During the review process, bursts can be used to detect abnormal activity. A professional reviewer will write new reviews for a variety of things over a long period of time,

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

from their first to their most recent. On the other side, some spammers who wish to sway the public's opinion against a specific product tend to write a huge number of reviews in quick span in order to swiftly outnumber honest evaluations. In this situation, their user identities are formed with the intent of writing spam reviews against one or more items, and they exhibit aberrant burst activity, which may be quantified using the model below. Equation(13) is used to compute the Reviewer burstiness of review's author a.

$$RBu(a) = \begin{cases} 0, & LR(a) - FR(a) > 30\ days \\ 1, & LR(a) - FR(a) \leq 30\ days \end{cases} \qquad (Eq.13)$$

where LR(a) represents author's recent and last review, while FR(a) indicates the date on which the author first post their review about any products they purchased through online. The above formulae represent that, if the author posted reviews have the difference in period of less than 30 days, then these reviewers are considered as a spammer. As a result, if an author total reviewing activity is limited over a long period of time, spamming suspicions may rise, and the Spamicity score of their reviews may suffer as a result.

Extreme ratings

Many spammers always use extreme ratings for the products through either high or low level to reach the goal of rapidly increasing or decreasing the average value of the products, respectively. It is observed that, a fake reviewer may provide a star rating of 1 instead of average star rating of 4 out of 5 star rating.

Let 'a' be the sum of the reviews given by Author 'A' which will help for the system to known about the individual behaviour towards calculating the EXR. The amount of Extreme Ratings (EXR) may be 1 or 5 is assigned for all the collected reviews and split based upon the grades represented by $|RS_a|$ which leads to the review's relation to extreme ratings and it is represented using the Equation (14).

$$EXR(a) = \frac{|RS_a \in \{1,5\}|}{|RS_a|} \qquad (Eq.14)$$

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

The above value contributes towards the identification of reputation of author *a*. A peak ratio with high score indicates that the reviewer is a spammer, but it must be further supported by other complementary spam indicators.

Reviewer average proliferation

An author has written reviews about other products earlier and this metric will help the suggested system to examining their review history, which is defined as a group of previous reviews given by the reviewer '*a'* for '*d'* discrete services or items, leaves a hint about the overall credibility or reputation of an author.The Author's Average Proliferation(AVgP) is calculated by using the Equation (15):

$$AVgP(a) = \frac{Size\ of\ Hist_{a,r}}{n} \qquad (Eq.15)$$

Where 'n' represents the number of products for which an author provided reviews in their history. Suppose if an author written 4 reviews for 4 products indicates non spamming activity, on the other hand if an author written 500 reviews (*SIZE of Hist*$_{a,r}$) for 4(n) products , then Author's Average Proliferation score is 125 which shows high spamming activity.

Spam Review Identification Metrics used for calculating the Review spamicity Score

Review can be classified as Genuine of fake based upon the reviews overall Spamicity Score S(r) and it is calculated based on the results obtained from the various metrics related to review and reviewer characteristics discussed earlier in previous sections.Scores obtained from each metric is multiplied by an appropriate value according to their importance and it is indicated in Table 2.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

Table 2. Various metrics affecting the Spamicity score, Purpose and its Weightage

Spamicity affecting Spam Metrics	Purpose	Weight
BuA(a)	Author Burstiness	1
RBu(a)	Review Burstiness	1
RD(r)	Rating Deviation	1
CF(r)	Classification Factor	0.5
EXR(a)	Extreme Rating	0.5
NR(a)	Number of Reviews	1
AvgP(a)	Average Proliferation	1
RRR(r)	Review Relevancy Rate	0.5
CL(r)	Content Length	0.5

Review Spamicity *S(r)* score is calculated by using the Equation (16).

$$S(r) = 0.5CF(r) + RD(r) + NR(a) + BuA(a) + RBu(a) + 0.5EXR(a) + AvgP(a) + 0.5RRR(r) + 0.5CL(r) \qquad (Eq.16)$$

For implemented system, It is decided to have a threshold value of 25 based on the conclusions of previous works for fake review identifications. Finally, if a review's score is larger than the cutoff value of 25 and it can be considered as a spam which is represented using Equation(17),

$$Review\ is\ consider\ as\ spam\ or\ honest = \begin{cases} Spam, S(r) > 25 \\ Honest, S(r) \leq 25 \end{cases} \qquad (Eq.17)$$

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

A compilation of all of the aforementioned metrics can be used to categorize the review as either spam or genuine as per the analysis results which includes the information related to review of content, context, creation time, burst behaviour, Proliferation and reviewer history of the individual author. When employing the suggested system to any kind of product or service, it will act as an effective filtering system which will removes fake or false content from the public space reviews.

PROCEDURE FOR COLLECTING THE DATA FOR THE PROPOSED SYSTEM

The evaluation of the SRIM is carried out via an experimental process, and for this purpose two discrete data sets are acquired. The data set selected here are from Kaggle website related to Amazon Mobile phone reviews. It contains a collection of 2,000 mobile phone reviews and they are split up across 4 categories: Positive and Genuine Reviews, Positive but Fake Reviews, Negative and also Genuine Reviews and Negative and Fake Reviews and these reviews are collected during their burst period of 7 days interval and it is represented through below Table 3.

https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

Table 3 MLP Training Dataset selection from the Amazon.com

Categories	Total number of Reviews	Number of Positive and Genuine Reviews	Number of Positive but Fake Reviews	Number of Negative and also Genuine Reviews	Number of Negative and Fake Reviews
Review Count	3000	750	750	750	750

The corpus consists of 3,20,47,028 reviews of 93,88,151 products, authored by 1,11,27,440 reviewers. The dataset consist of attributes like creation date, rating, and review text. The original dataset may include some noise data and it should be pre-processed before further processing. After preprocessing step (like tokenization and removal of stop words), the final data set consists of 2,89,74,127 reviews of 87,12,787 products of 1,01,12,174 reviewers.

EXPERIMENTATION METHODOLOGY

Evaluation methods

For any research oriented activity, it is necessary to find out the effectiveness of the proposed system over existing systems. It is difficult to find an exact method to compare the performance of implemented techniques with the results obtained from the existing systems. Previously human intervention is made to compare the performance between the systems.

The suggested method assesses all kinds of products or services based on reviews, awarding a score to each reviews to calculate the spamicity. Text categorization is employed in assessing reviews and it faces the difficulties like owing to memory restrictions, classification is

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

particularly difficult for the entire trial data set. As a consequence, to make the text classification assessment approach easier to use, the suggested system will divide the entire reviews in the three sub-parts are sorted in decreasing order according to their spam score. In other research, the top K rankings correspond to the positive class related to spam reviews , whereas the bottom K ranks correspond to the negative class related to honest reviews. For each of the above class will have 1,500 reviews, which is a small enough number to allow for execution yet large enough to offer a representative sample size. After the written information from the 3000 reviews has been obtained, it is recovered and used in the rating evaluation process. Finally, performance reporting and 5-fold cross-validation are utilised to evaluate the classifier's accuracy and, as a result, the spam detection system's overall efficacy.

The Multilayer Perceptron technique is assessed separately and used as a parameter for calculating the spamicity score, that is used to categorise review content as part of the wider spam detection system. Using the gold standard dataset of annotated reviews and 10-fold cross-validation, the classification is compared against several baseline methodologies specialised in review text categorization and reporting of the resulting accuracy scores. The next part goes through the accuracy measurements used in the recommended studies.

Performance Evaluation Metrics

Precision and recall measures are used to quantify the accuracy and efficiency of the recommended technique. These metrics are used in previous studies and these are used to measure the accuracy of review spam detection. Precision refers to the proportion of spam reviews among all really deceptive reviews, whereas Recall refers to the percentage of fake reviews among all truly deceptive reviews. These two computations can be summarized using Equations (18) and (19).

$$Precision = \frac{|\{Spam\ Reviews\} \cap \{Detected\ Reviews\}}{\{Detected\ Reviews\}} \qquad (Eq. 18)$$

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

$$Recall = \frac{|\{Spam\ Reviews\} \cap \{Detected\ Reviews\}}{\{Spam\ Reviews\}} \qquad (Eq.19)$$

Accuracy is measured using the Equation (20).

$$Accuracy = \left(\frac{Number\ of\ reviews\ correctly\ classified\ as\ spam\ by\ \ SRIM}{Total\ number\ of\ actual\ spam\ Reviews\ in\ dataset}\right) (Eq.20)$$

RESULTS AND DISCUSSIONS

Evaluation and Parameter Selection Procedure for Review Content Classification

Using n-gram characteristics, the MLP Classifier will evaluate the review content as spam or honest. Frequently used classifiers like Nave Bayes and SVM have been used in the literature, and the accuracy has been measured using a gold standard set for mobile phone datasets. The results will give a relevant comparison with other existing systems because the system uses a machine learning approach and is tested with error detection tools. To determine the MLP classifier's behavior in each category, accuracy values for positive, negative, and mixed opinions are presented.

Furthermore, the dimensions of the vector representation in this problem are based on n-grams, but the number of neurons is typically connected to the number of input nodes. In order to eliminate unnecessary dimensions, Principle Component Analysis (PCA) uses an orthogonal transformation to change a collection of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of main components is fewer than or equal to the number of original variables or observations. Table 4 displays the experimental results for the number of neurons in the hidden layer of the MLP classifier.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

Table 4. Experimental results to calculate the accuracy

Number of neurons	Precision	Recall	F-Measure	Accuracy%
300	0.846	0.84	0.832	84.10%
600	0.881	0.872	0.871	87.55%
800	0.841	0.846	0.841	84.25%
1000	0.834	0.825	0.823	82.40%

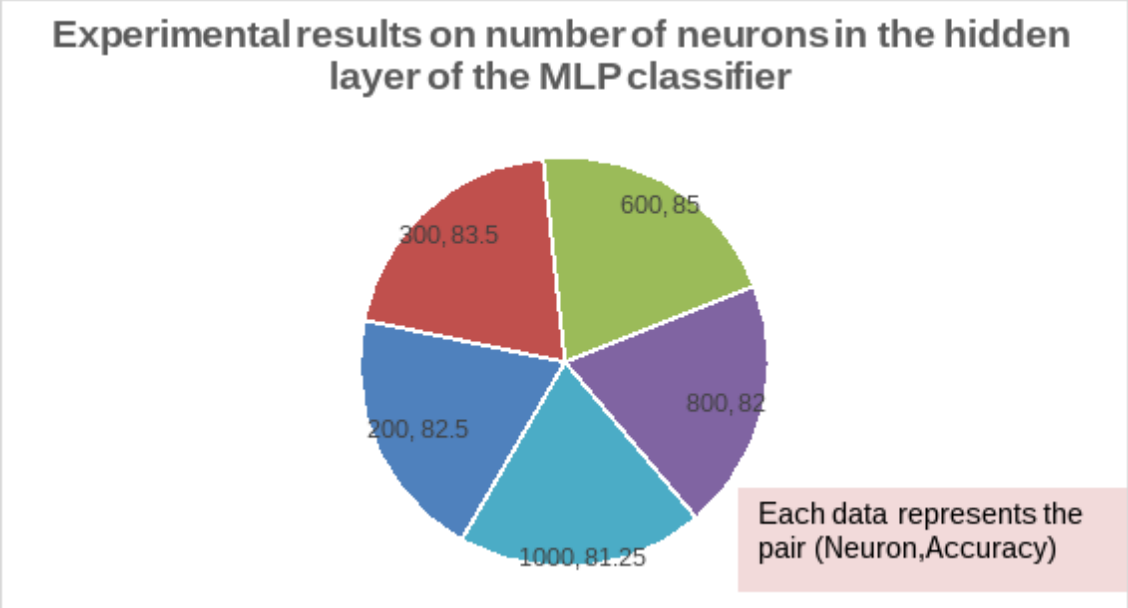

Figure 4. Experimental results of accuracy of proposed system

From Figure 4, the results concluded that, Neuron size(*N*) of 600 produced high accuracy than the other sample size when the Epoch value of 10 is mentioned in Table 4.

The evaluation of accuracy by the proposed methodology is based on Positive reviews, Negative Reviews and Combined reviews of the Amazon review dataset. The overall accuracy of the MLP classifier are presented in below Table 5.

https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

Table 5. Accuracy comparison of MLP classifier with NB and DT Techniques

Train Sentiment	Precision	Recall	F-Measure	MLP Accuracy	Naïve Bayesian Classifier Accuracy	Decision Tree Accuracy
Positive(1000 reviews)	0.862	0.862	0.862	86.25%	88.53%	86.70%
Negative(1000 reviews)	0.866	0.865	0.864	86.5%	84.14%	82.31%
Combined(2000 reviews)	0.856	0.855	0.855	85.56%	84.02%	81.18%

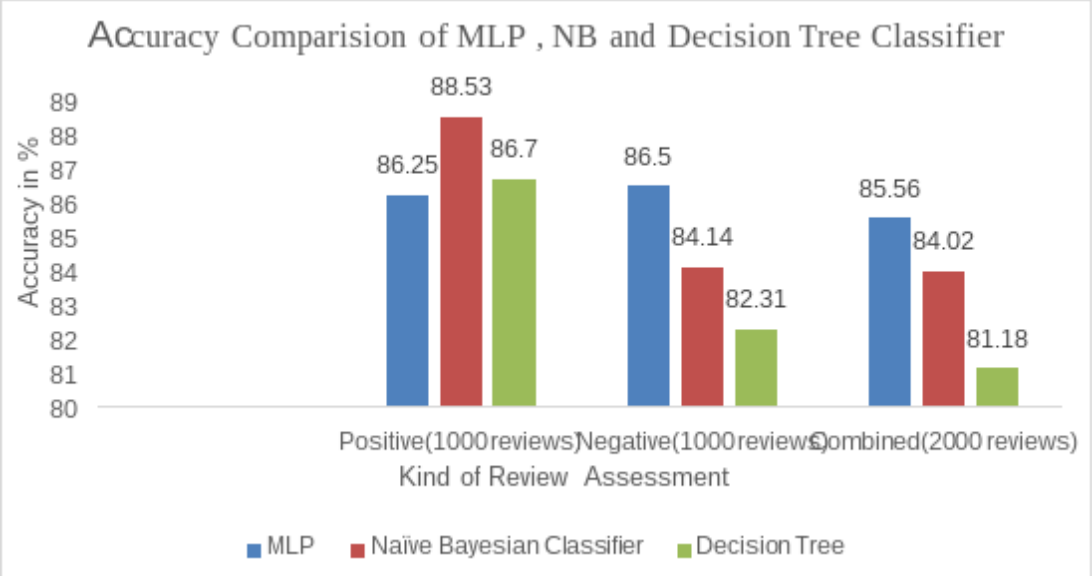

Figure.5 Accuracy Comparison of MLP vs NB vs DT Classifier

An interesting observation from Figure 5 states that, although the positive sentiment performance with respect to NB and Decision Tree outperforms well by a accuracy difference of 2.28% and 1.45% than MLP classifier, however MLP produced good results for negative sentiment related reviews by a accuracy difference of 2.36% and 4.37 % than NB and DT methods. MLP also produced good accuracy improvement for combined reviews than NB

https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

and DT methods since Multilayer Perceptron works well with negative sentiment word and combined review rather than the positive kind of words. Since many organizations appointed spammer to write negative reviews against their competitors than positive kind of reviews.

Detection of spam reviews using SRIM through Spamicity Score

Table 6 shows the scoring approach used to arrive at the spamicity scores for the top rated reviews in the dataset.

Table 6. Spam scores calculation for the highest ranked reviews present in the entire dataset

CF	*RD*	*CL*	*NR*	*BuA*	*RBu*	*RRR*	*EXR*	*AvgP*	Spam Score
0	0.63	1	108	0	0	52	0.612	2.31	69.64
1	0.26	1	2	1	0	44	1	57	43.01
1	0.03	1	25	1	1	41	1	16.25	35.96
0	0.02	1	29	1	1	39	1	13	34.12
1	0.12	1	18	1	1	42	1	17	33.79
1	0.47	1	36	0	1	25	0.98	1	29.36
0	0.01	1	2	1	0	41	0.992	31	28.51
1	0.21	1	31	0	1	29	0.64	1.5	27.28
1	0	0	27	0	1	32	1	1	25.96
0	0.28	0	15	1	1	34	1	8	25.28

From the findings of Figure 6 , a few interesting conclusions can be drawn. It is obvious that Content Length (CL), Review Relevancy Rate (RRR), and Number of Reviews (NR) are quite valuable and operate as an obvious indicator in analysing reviewer behaviour, which is then used to determine if a review is spam or honest. Those reviews whose authors provide a large number of reviews for a single product receive very high spamicity scores. The findings also show that review relevance is high among the reviews, since spammers either duplicated or modified existing review content to create comments for other goods using their

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

own or alternative credentials. The reviews are also written in multiple lines (Content length) rather than a single line of text by the spammer.

Furthermore, while some research has focused solely on bursty times as a means of detecting spam activity, it is clear that this method has the shortcoming of ignoring reviews that is not written with in a particular burst period. The above table indicates that more than one instance of a review in which the reviewer was not in the burst peroid, resulting in CSBu and BuA values of zero. Aside from the individual evaluations provided by the RBu and EXR parameters, the contribution of author reputation is represented using reviewer average proliferation on previous history. Three occurrences of The value of AvgP reaches maximum at three instances out of which one author had a maximum NR value for a product and rest two for singleton reviews.

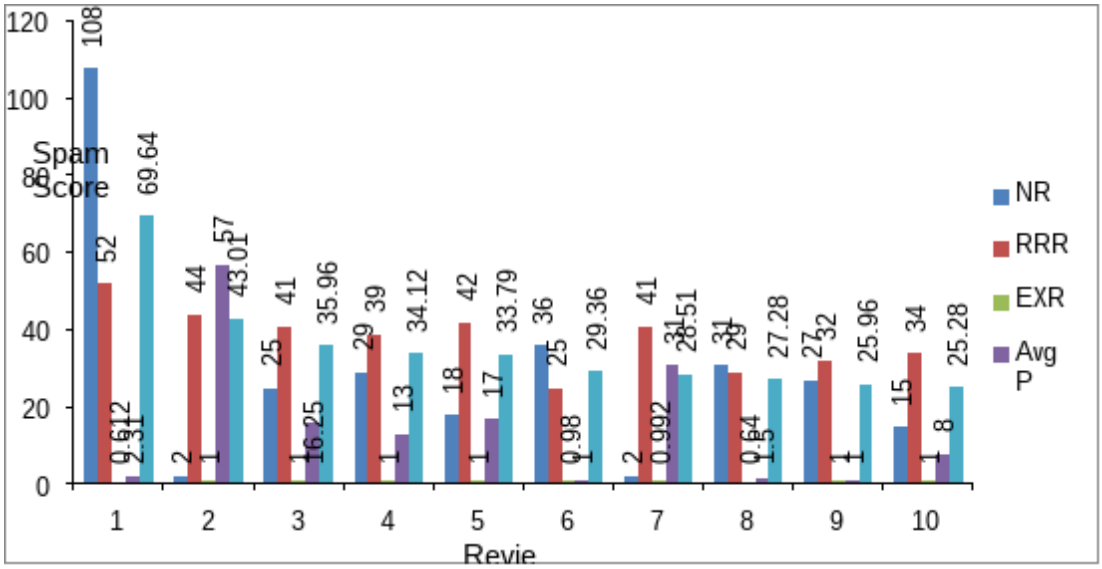

Figure.6 Spam scores calculation for the highest ranked reviews present in the entire dataset

https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

CONCLUSION

SRIM focused on the opinion spam detection within online reviews and it employed several metrics to detect effectively spam reviews rely on various metrics related to reviews and reviewer behaviour and their reputation. There are various malicious entity often attempts to exploit the review sentence to improve or degrade the products or services due to business competition.

MLP classifier is used to determine the usefulness of the review by categorizing the review as spam or honest. Negative based reviews are identified with improved accuracy over the existing classifier like Naïve Bayes and SVM by a difference of 2.36% to 4.37% respectively since many users tends to write negative reviews about their competitors' products. So, it is essential to determine that, whether the negative reviews are written by either genuine user or spammers.

The experimental and evaluation phase of the spam review detection based on the content and context is presented in the previous section. The implemented system used for labelled the review as either spam or honest as the result of spamicity function given an indication about the review . These are major spam instances since they involve a large number of reviews for a single product written by the same author. In fact, the availability of singleton spam reviews increases the likelihood of spam reviews, as 92 percent of reviewers react on singleton reviews.

Below are the lists of inferences from the proposed system,

- Proposed experiment included the detection of the burst pattern (time window), since it plays a major role to detect the time taken by the spammer to write the fake reviews. Burst pattern detection technique is used for testing on different values based on time window parameter, and it is evaluated that the 7 days are the most accurate and appropriate value spent by the spammer rather than 14 or 30 days detected by previous

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

studies. Based on burst pattern detection, the values for the rest of the metrics are calculated towards Spamicity Score.

- Implemented experiment also consider the author's reputation to detect the spam reviews. Users with widely past spam reviewing activity (high AvgP values) are identified and filtered their reviews from the original dataset.

- Implemented system also indicate that,some spammers looks alike the real reviwer which is considered by the current methodology into account for spam detection and thus increases efficiency.

- As a result, the evaluation for SRIM for filtering the review dataset based on text classification is performed and it yields positive and reasonable results. Given the evaluations made from the top rated review , then the overall accuracy hits almost 81.5% due to the availability of more spam activity conditions that have been successfully captured with the proposed methodology. While more reviews and reviewers are expected to offer more behavioural characteristics to analyse, more than 80% accuracy has been reported for products with more review activity (Most Popular Products).

REFERENCES

1. A. Bagheri, M. Saraee & F. De Jong 2013, ‘Care more about customers: Unsupervised domain-independent aspect detection for sentiment analysis of customer reviews’, Knowledge-Based Systems, Vol. 52, pp. 201–213.
2. Adnan Duric, Fei Song (2012), ‘Feature selection for sentiment analysis based on content and syntax models’, Decision Support Systems, Vol. 53, No. 4,Pp. 704-711
3. Chaowei Zhanga, Ashish Guptaa, Christian Kautena ,Amit V.Deokarb , Xiao Qin 2019, ‘Detecting fake news for reducing misinformation risks using analytics approaches’,European Journal of Operational Research, Science Direct, Vol. 279, No. 3, pp. 1036-1052.
4. Chih-Chien Wang, Min-Yuh Day & Yu-Ruei Lin 2016, ‘Toward Understanding the Cliques of Opinion Spammers with Social Network Analysis’, International Conference on Advances in Social Networks Analysis and Mining, pp. 1163-1169.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

5. Dinesh Kumar Vishwakarma, Deepika Varshney ,Ashima Yadav 2019, 'Detection and veracity analysis of fake news via scrapping and authenticating the web search', Cognitive Systems Research, Science Direct ,Vol. 58, pp 217-229.

6. Eka Dyar Wahyuni & Arif Djunaidy 2016, 'Fake Review Detection from A Product Review Using Modified Method of Iterative Computation Framework', MATEC Web of Conferences, vol. 58

7. Fei, G., Mukherjee, A., Liu, B., Hsu, M., Castellanos, M., & Ghosh, R. 2013, 'Exploiting burstiness in reviews for review spammer detection', International Conference on Web logs and Social Media, pp.55-105.

8. Huayi Li, Zhiyuan Chen, Bing Liu, Xiaokei Wei & Jidong Shao 2014, 'Spotting Fake Reviews via Collective Positive-Unlabeled Learning', IEEE International Conference on Data Mining, pp. 899-904.

9. Jindal, N & Liu, B 2008, 'Opinion Spam and Analysis', International Conference on Web Search and Web Data Mining, pp. 219–230.

10. Julien Fontanarava, Gabriella Pasi & Marco Viviani 2017, 'Feature Analysis for Fake Review Detection through Supervised Classification', International Conference on Data Science and Advanced Analytics, pp. 56.

11. J. Malbon 2013, 'Taking fake online consumer reviews seriously', Journal of Consumer Policy, Vol. 36, no.2 pp. 139-157.

12. J. Yi, T. Nasukawa, R. Bunescu & W. Niblack 2003, 'Sentiment analyzer: extracting sentiments about a given topic using natural language processing techniques', Third IEEE International Conference on Data Mining, USA, pp. 427-434.

13. K. Zhang, R. Narayanan & A. Choudhary 2009, 'Mining online customer reviews for ranking products', EECS Department, North-western University.

14. K. Zhang, R. Narayanan & A. Choudhary 2010, 'Voice of the customers: Mining online customer reviews for product feature-based ranking', Wonference on Online social networks, USA.

15. Lalit Pathak, Sanjana Mulchandani, Mayuri Kate, Kajal Khatwani & Sujata Khedkar 2018, 'Sentiment and Intent Analysis for Business Intelligence', International Journal of Innovative Research in Computer and Communication Engineering, Vol. 6, no. 1, pp.348-354.

16. Manali S.Patil & A.M.Bagade 2012, 'Online Review Spam Detection using Language Model and Feature Selection', International Journal of Computer Applications, Vol. 59, no.7,pp.33-36.

17. Mukherjee, A., Venkataraman, V., Liu, B., & Glance, N. 2013, 'Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews', pp.35-55.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

18. M. Hu & B. Liu 2004, 'Mining and summarizing customer reviews', International Conference on Knowledge, Discovery and Data Mining, ACM, Vol. 4, pp. 168-177.

19. M. K. Dalal & M. A. Zaveri 2014, 'Opinion Mining from Online User Reviews Using Fuzzy Linguistic Hedges', Applied Computational Intelligence and Soft Computing, Vol. 2014, no. 1, pp. 1–9.

20. M. N Istiaq Ahsam, Tamzid Nahian, Abdullah All Kafi, Ismail Hossain & Faisal Muhammad Shah 2016, 'Review Spam Detection using Active Learning', Annual Information Technology, Electronics and Mobile Communication Conference, IEEE, pp. 1-7.

21. Ott, M., Cardie & C., Hancock, J.T 2013, 'Negative Deceptive Opinion Spam', Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, USA, pp. 497-501

22. Paroubek P 2007, 'Evaluating Part-of-Speech Tagging and Parsing Patrick Paroubek, Speech and Language Technology', Springer, Vol. 37, no.9, pp.35-55.

23. P. Balaji, O. Nagaraju & D. Haritha 2017, 'Levels of sentiment analysis and its challenges: A literature review', International Conference on Big Data Analytics and Computational Intelligence, pp. 436-439.

24. Ravi Kumar V & K. Raghuveer 2013, 'Dependency driven semantic approach to Product Features Extraction and Summarization Using Customer Reviews', vol.178 pp. 225–238.

25. Sonu Liza Christopher & Rahulnath H A 2016, 'Review Authenticity verification using supervised learning and reviewer personality traits', International Conference on Emerging Technological Trends, pp. 1-7.

26. T. Hennig-Thurau, K.P. Gwinner, G. Walsh, & D.D. Gremler 2018, 'Electronic word-of-mouth via consumer-opinion platforms: what motivates consumers to articulate themselves', International Journal Interaction and Marketing, vol.18, no.1pp.38–52.

27. Wang, S. Zhu, & T. Li, 'SumView: A Web-based engine for summarizing product reviews and customer opinions', Expert System Applications, Vol. 40, no. 1, pp. 27–33.

28. W. Zhang, H. Xu, & W. Wan 2012, 'Weakness Finder: Find product weakness from Chinese reviews by using aspects-based sentiment analysis', Expert System Applications, Vol. 39, no. 11, pp. 10283–10291.

29. Xie S., Wang, G., Lin, S., & Yu, P. S. 2012, 'Review spam detection via temporal pattern discovery', International conference on knowledge discovery and data mining, ACM, pp. 823–831.

30. Yuan Yuan, Sihong Xie, Chun-Ta Lu, Jie Tang & Philip S. Yu 2016, 'Interpretable and Effective Opinion Spam Detection via Temporal Patterns Mining across Websites', IEEE International on Big Data, pp. 96-105.

 https://doi.org/10.37896/pd91.5/91538

ISSN: 0369-8963

31. Yuming Lin, Tao Zhu, Hao Wu, Jingwei Zhang, Xiaoling Wang & Aoying Zhou 2014, 'Towards Online Anti-Opinion Spam: Spotting Fake Reviews from the Review Sequence', International Conference on Advances in Social Networks Analysis and Mining, pp. 261-264.

32. Zhang L., Liu B. 2017, 'Sentiment Analysis and Opinion Mining', Encyclopedia of Machine Learning and Data Mining. Springer, Boston, MA, pp. 1152 - 1161.

 https://doi.org/10.37896/pd91.5/91538