International Journal of Science and Research (IJSR)
ISSN: 2319-7064
ResearchGate Impact Factor (2018): 0.28 | SJIF (2018): 7.426

A Novel Approach for Genuinity Analysis of Hotel Online Reviews

Pankaj Chauhdary[1], Dr. Anurag Aeron[2], Dr. Sandeep Vijay[3]

[1]Research Scholar, ICFAI University, Dehradun, India

[2]Associate Professor (CSE), ICFAI University, Dehradun , India

[3]Director, Shivalik College of Engineering, Dehradun, India

Abstract: ***Background**: Since previous decades Internet as well as smart phones have become easily accessible to maximum people. This has made social networking an integral part of human life. People are sharing their comments and reviews on the forum or portal about their views and experiences. These reviews help others to judge the brand value of any product. Even in taking the final decisions about the brand selections for best hotels, colleges and products people are gradually depending on the previous online reviews. In such scenario, some companies may indulge themselves in generating the fake reviews with wrong intentions to create the positive or negative hype about the particular products. It may mislead the customers and decision makers. **Objectives**: Objective is to develop an algorithm to development of the optimal machine learning algorithm for hotel reviews Efforts are made to remove maximum limitations and constraints of existing algorithms to develop a robust algorithm. **Methodology**: After finding the gaps appropriate mathematical models are proposed to be implemented to detect genuinety of the reviews based on behavior metrics, quantify the past trust analysis of the reviewer, group membership activities and quantify the sentimental analysis for the hotels. **Findings**: Due to filtration of the spam reviews and fake reviewers, systematic predication about the hotel facilities and ambience may be done that will encourage the customer to use the hotel booking website that will utilize such algorithms. **Applications/Improvements**: Although this work is specifically proposed for helping customers in selection of the best hotels by analyzing the previous online reviews, and help in concluding the right decision based on Location, Security, Price, Quality, Ambiance etc. Yet the something similar model may be designed after minor modifications for taking right decision in selecting the best colleges, best products etc.*

Keywords: Classification, Machine learning, Burst rate, sentimental analysis, past trust analysis etc

1. Introduction

In current ear, if maximum hotel bookings are online. In case of hotels more positive reviews earn more reservations and business. It can be tempting to request friends, family, and employees to leave positive reviews online for the hotels or even to pay for high marks online. However, aside from being unethical and misleading, fake reviews can have serious consequences. Supposed we want to travel abroad, Fake reviews can literally spoil our travelling experience in a new country [1].

1.1 Behaviour matrices

Eight mathematical characteristics for unusual behaviour[2]. of data sets. With some modification, we may propose following quantified indicators for hotel reviewers.

a) Customer priority
b) Deviation rate
c) Bias rate
d) Review Similarity rate
e) Review Quality Relevance
f) Content Length
g) Illustration.
h) Burst rate

1.2 Past trust analysis

Once social relationship is properly identified using a graph[3]. Individual user can be assessed based on following parameters that are available in public domain also. These are Reviews generated by the user in past, Ratings provided, Photos uploaded , Videos uploaded , Answers, Edits , Places added, Roads added, Facts Checked, Q&A.

1.3 Analyze the group membership activities

Group Membership and Social Influence

The Social influence and association among various reviewers plays a major role. Structural social psychology theories illustrate how the group or the network structures may seriously affect the individual outcomes e.g. to exchange profits, self-identities, locations within the hierarchies. It may happen that some individuals may not be aware sometime to the source of influence. Nor able to recognize and respond to relatively unknown factors, such as threat was posed by unidentified outsiders group but that is real in actual [4,5,6,7].

In such cases, impact of factors due to association may provide an accurate understanding of the behaviours, experiences and consequences [8]. In day to day life people may make false inferences for others based on observable characteristics without having much knowledge about task-relevant abilities of others.

1.4 Self Categorization theory

Self categorization theory was defined based on the concept of formations of psychological group. This theory specially emphasizes on categorization processes. With the help of cognitive underpinnings. It concludes that the process of group categorization results in depersonalization. Group members are interchangeable [9,10]. According to this theory the people usually may establish confidence in their opinions by comparing the beliefs provided by similar psychological group members.

1.5 Status characteristics theory

For a particular person, A status characteristic can be defined as a property that may be assigned two or more states or levels with separate values, each state or level usually is associated with one or more similarly evaluated expectations. Higher status members are those members that are advantaged with respect to the group's observable power and prestige order (OPPO) [12,14].

Those actors, with higher status have following properties

a) They are provided more opportunities to make suggestions in the group decisions.
b) Usually it is assumed that their suggestions are relatively better.
c) Maximum suggestions provided by them are positive suggestions
d) Their suggestion is more robust and have more influence over other members' opinions.

The status characteristics theory is applied with a purpose

a) To solve a group task by considering other's suggestions.
b) Consideration of both correct and incorrect solutions is necessary to solve the task .

The theory consists of five characteristics:

a) **Salience:** A member will be considered as salient If it perceived as relevant to the task, and status characteristic can easily differentiate the members.
b) **Burden of proof:** When status characteristic is salient and task has not been disassociated, expectations consistent with states of the characteristic are formed by the actor.
c) **Sequencing:** If actors ensure exit of enter on tasks to perform the expected tasks performance, status information and sequencing is preserved.
d) **Combining:** To form aggregated expectation sets, the effects of multiple similarly evaluated status characteristics may is combined combined.
e) **Basic expectation assumption:** If a person is dependent on expectations to infer competence, then the better competence results will come with greater person's in the person's higher position [16,18].

1.6 Proposed model for analyzing group membership activities in hotel reviews

Let U is set of possible users, R is the set of rating and P represent the set of products respectively in a graph G=(U,R,P). Supposed user u € U assign a rating(u,p) € R to the product p € P. We assume that rating scores are approximated between -1 and +1. Users in terms of their fairness or trustworthiness may vary. Fair users without bias usually give good scores to good products and bad products are assigned low scores. On the other hand, fraudulent users with wrong intentions assign bogus high ratings to low quality products and low ratings to the good products. For a user u, Fairness score F (u) may be identified by analyzing the ratings by all members of the group and it lies in the [0, 1] interval ∀u ∈ U. Here 0 will denote the untrustworthy user, whereas 1 denotes fully trustworthy user [19].

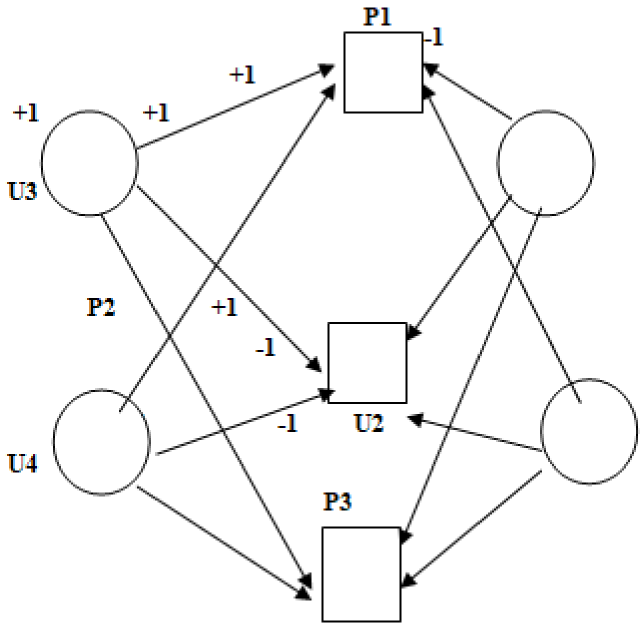


Figure 1: Group theory implementation

Above theory may be easily implemented in hotel reviews. e.g. In above example U1 is fake user.

2. Quantify the Sentimental Analysis

2.1 Opinion Mining

Sentiment Analysis (SA) or Opinion mining, is the process to analyze people's opinions, appraisals, sentiments, evaluations, attitudes, and emotions towards entities such as services, topics, individuals, issues, and their attribute. It is formulated as a two-class classification problem, positive and negative. Sentimental Analysis is the process of analysing the positive or negative polarity of a given text at three levels i.e. **document, sentence** or **aspect level** [21].

2.2 Textual reviews

To analyze the textual reviews reputation models depend on numeric data available in different fields that is derived based on the consumers textual reviews to provide a detailed opinion about the product. With changing time customers are giving more importance to the reviews rather than the numeric ratings.

2.3 Sentiment analysis issues

Majorly two major issues are encountered while considering Sentimental Analysis. First, the opinion observed as negative in some situation might be considered as positive in other situation. Second, people may not always express opinions in the similar way [23,24,25].

2.4 Detecting Fake Reviews Using Machine Learning

Several machine learning algorithms such as supervised, unsupervised, semi supervised and re-enforcement learning may be utilized for sentiment classification at document level for declaring a negative or positive sentiment. A confusion matrix is generated to classify the review as positive and negative. Following terms are used in quantification.

True Positive: True positive(TP) reviews are that reviews that are correctly classified by the classification model as positive .

False Positive: False Positive (FP) are that reviews that are wrongly classified as Positive by the classification.

True Negative: The reviews that are correctly classified as Negative by the classification model are termed as True Negative (TN).

False Negative: The reviews that are incorrectly classified as Negative by the classification model are termed as False Negative (FN).

$$\text{Fake Positive Review Rate} = \frac{FP}{FP+TN}$$

$$\text{Fake Negative Review Rate} = \frac{FN}{TP+FN}$$

$$\text{Real Positive Review Rate} = \frac{TP}{TP + FN}$$

$$\text{Real Negative Review Rate} = \frac{TN}{TN+FP}$$

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FN+FP}$$

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

3. Proposed Algorithm

Based on the overall work, discussions and hypothesis justifications following algorithm is derived.

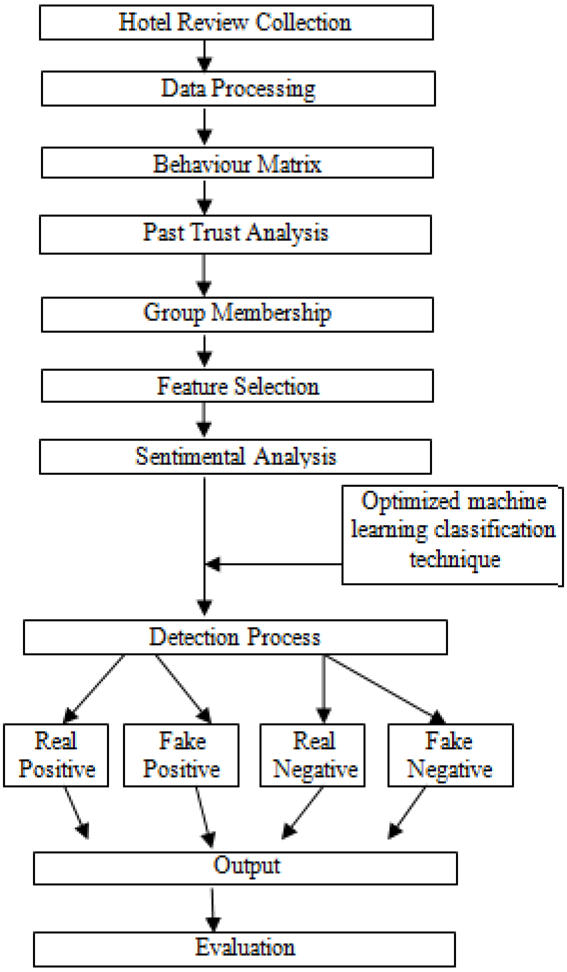


Figure 2: Review genuinity analysis algorithm

4. Analysis and Results

A detailed survey was conducted on 602 Reviews of Hotel Grand Legacy given in grand legacy and following parameters are suggested as key indicators of behaviour metrics: and following findings were there.

Table 1: Behavior Matrices

SN	Parameters	Genuine Reviews
1	Percentage Reviews that passed Customer priority test	76%
2	Percentage Reviews that passed Review Similarity rate test	82%
3	Percentage Reviews that passed Review Quality Relevance test	83%
4	Percentage Reviews that passed Content-Length test	78%
5	Percentage Reviews that passed Illustration test	84%
6	Percentage Reviews that passed Burst Rate test	89%

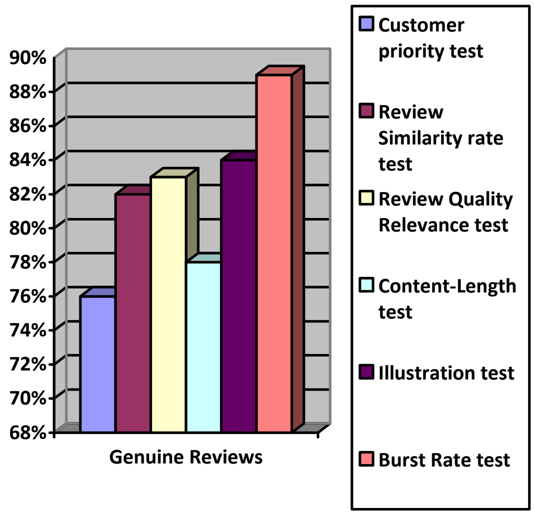


Figure 2: Genuine reviews passed behavior matrix tests

Table 2: Past Trust Analysis

SN	Test Parameters	Test passed by Reviews
1	Reviews	92%
2	Ratings	93%
3	Photos	97%
4	Videos	96%
5	Answers	94%
6	Edits	93%
8	Places added	91%
9	Roads added	92%
10	Facts Checked	90%
11	Q&A	89%

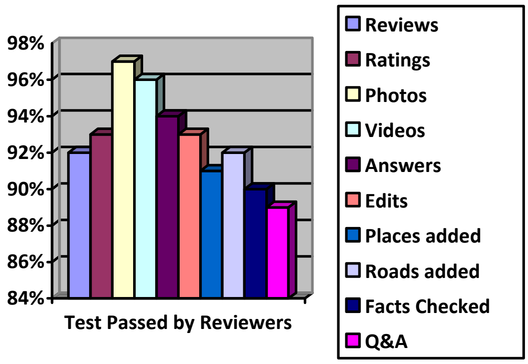
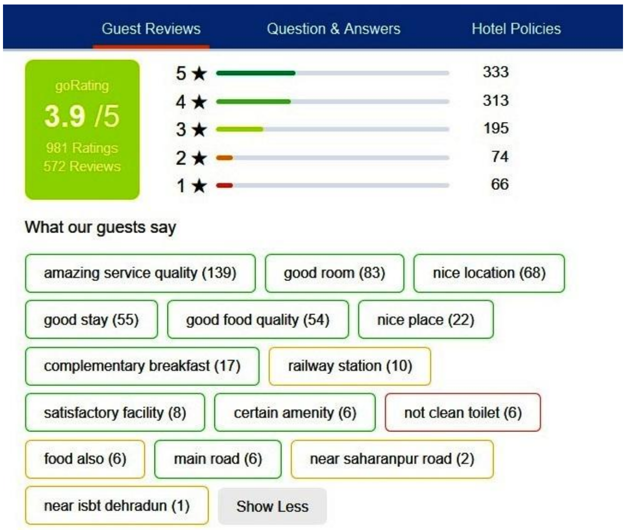


Figure 3: Past trust analysis

Table 3: Sentimental Analysis and Genuine reviews

SN	Keywords for Machine Learning Classification	Agreed by Reviews
1	Amazing service quality	92%
2	Good room	93%
3	Nice Location	97%
4	Good Stay	96%
5	Good Food Quality	94%
6	Nice place	93%
8	Complimentary breakfast	100%
9	Railway station	98%
10	Satisfactory facilities	73%
11	Certain amenity	79%

4.1 Advertised Reviews and rating



4.2 Suggested reviews and rating as per the algorithm

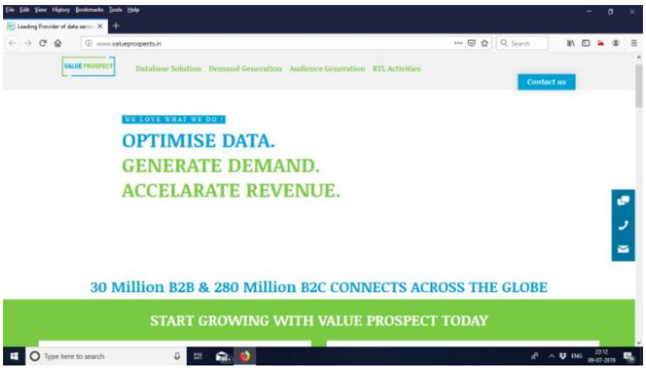
After dropping the less important reviews that failed to pass the various tests. Following is the conclusion.

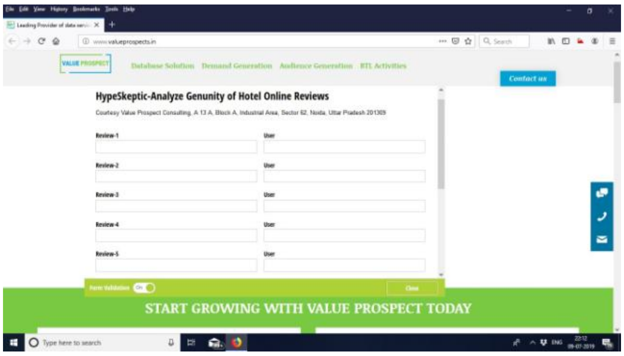
Table 4: Genuine reviews

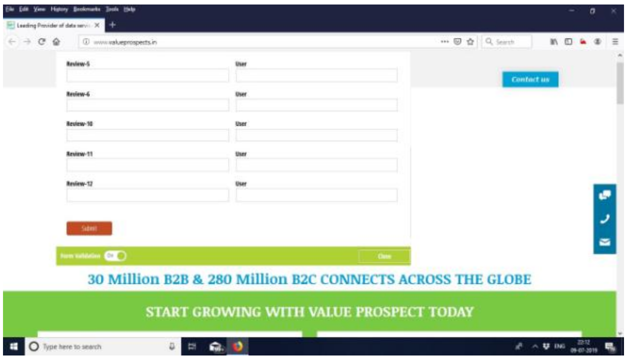
SN	Keywords for Machine Learning Classification	Agreed by Genuine Reviews
1	Amazing service quality	106
2	Good room	71
3	Nice Location	48
4	Good Stay	39
5	Good Food Quality	42
6	Nice place	14
8	Complimentary breakfast	14
9	Railway station	8
10	Satisfactory facilities	1
11	Certain amenity	1

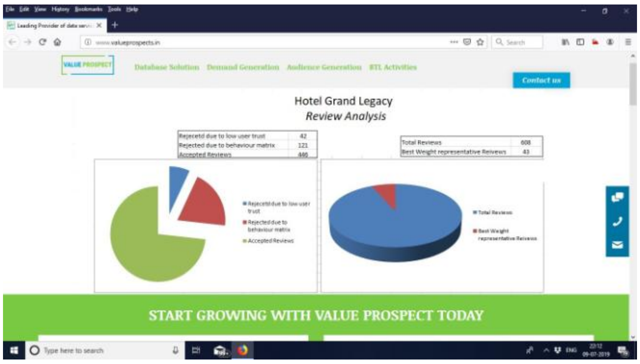
Therefore we may conclude that hotel do not have good service quality but breakfast is complementary and it is near to railway station, location is average. And overall rating of the hotel is **3.4/5**

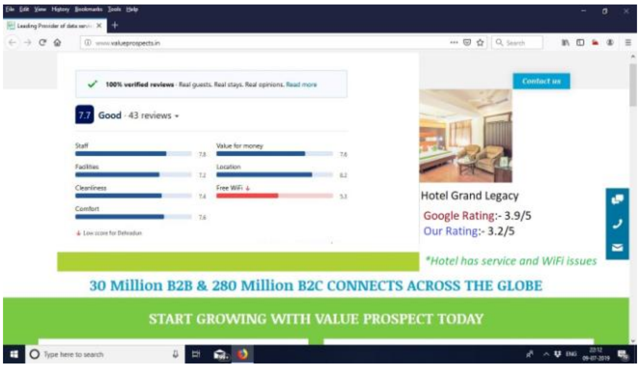
4.3 Web interfaces of the proposed tool











5. Conclusion and Future Work

Above methods of quantification of genuine reviews are based on mathematical models and can give better results as well as less important to the fake reviews also to the fake reviewers may further be ignored while calculate genuine conclusion about the parameters of the hotels. Our model can further be improved by mathematically improving the procedures to calculate Customer priority, Deviation rate, Bias rate, Review Similarity rate, Review Quality Relevance, Content Length, Illustration, Burst rate. Also web regulators may be requested to provide more information publically about the activities of the reviewers in addition to Reviews given, Ratings provied, Photos uploaded, Videos, Answers, Edits, Places added, Roads added, Facts Checked, Q&A. Sentimental analysis may also be further improved to ensure robust classification as per appropriate machine learning technique.

References

[1] Visani C. and Jadeja N., (2017), "A Study on Different Machine Learning Techniques for Spam Review Detection", IEEE Transaction 978-1-5386, pp. 1887-1892.

[2] Xue H., Li F., Seo H. and Pluretti R., (2015), "Trust-Aware Review Spam Detection", IEEE Computer Society Trustcom/BigDataSE/ISPA, pp. 726-733.

[3] Jiang M. and Cui P., (2016),"Suspicious Behaviour Detection: Current Trends and Future Directions", IEEE Intelligent systems/1541-1672/16, Computer Society, pp. 31-39.

[4] Rout J.K., Dalmia A. and Choo K. K. R., (2017), "Revisiting Semi-supervised Learning for Online Deceptive Review Detection", IEEE Access, 2169-3536, pp. 11-19.

[5] Deng X. and Chen R., (2014), "Sentiment Analysis Based Online Restaurants Fake Reviews Hype Detection", proceedings of APWeb Workshops, Springer International Publishing Switzerland, pp. 1–10.

[6] Ruchansky N., Seo S. And Liu Y., (2017), "A Hybrid Deep Model for Fake News Detection", Computer Society of India ACM. 978-1-4503-4918-5/17/11, pp. 17- 27.

[7] Mukherjee A., Venkataraman V., Liu B., Glance N., (2013), "Fake Review Detection: Classification and Analysis of Real and Pseudo Reviews", Technical Report, Department of Computer Science (UIC-CS-2013-03).University of Illinois at Chicago, pp. 268-279.

[8] Kokate S., Tidke B., (2015),"Fake Review and Brand Spam Detection using J48 Classifier", International Journal of Computer Science and Information Technologies ISSN: 0975-9646, Vol. 6 (4), pp. 3523-3526.

[9] Chaitanya Kale, Dadasaheb Jadhav., Tushar Pawar., (2016), "Fake Spam review detection using natural language processing techniques", International journal of innovations engineering research and technology ISSN: 2394-3696, Vol. 3, Issue 1, Jan.-2016, pp. 31-37.

[10] Adike R.G., Reddy V., (2016), "Detection of Fake Review and Brand Spam Using Data Mining Technique", International Journal of Recent Trends in Engineering & Research (IJRTER) Volume 02, Issue 07; July - 2016 [ISSN: 2455-1457], pp. 251-256.

[11] Bonde Y.P., Kharabi K.L., Sabale A.N., (2017), "Detection and Elimination of Fake Review from Real-Time Data using Cloud Computing", International Journal of Advance Engineering and Research Development ISSN: 2348-6406 Volume 4, Issue 5, May-2017, pp. 187-194.

[12] Crawford M., Khoshgoftaar T.M., Prusa J.D., Richter A.N. and Najadain H.A., (2015), "Survey of review spam detection using machine learning techniques", Springer Journal of Big Data Machine Learning Methods Crawford et al, pp. 17- 39.
[13] Elmurngi E. and Gherbi A.,(2017),"An Empirical Study on Detecting Fake Reviews", proceeding of IEEE The Seventh International Conference on Innovative Computing Technology, pp. 107-114.
[14] Fontanarava J., Pasi G. and Viviani M., (2017), "Feature Analysis for Fake Review Detection through Supervised Classification**",** proceedings of IEEE International Conference on Data Science and Advanced Analytics, pp. 658-666.
[15] Wahyuni E. D. And Djunaidy A., (2017), "Fake review detection from a product review using modified method of iterative computation framework", proceedings of MATEC Web of Conferences, pp. 121-127.
[16] Lin Y., Zhu T., WuI H., Zhangl J., Wang X., Zhou A., (2014), "Towards Online Anti-Opinion Spam: Spotting Fake Reviews from the Review Sequence", proceedings of IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, pp. 261-264.
[17] Li Y., Feng X., Zhang S., Li Y., (2016), "Detecting Fake Reviews Utilizing Semantic and Emotion Model", proceedings of 3rd IEEE- International Conference on Information Science and Control Engineering, pp. 317-320.
[18] Yin R., Wang H., Liu L. , (2015), "Research of Integrated Algorithm", proceedings of 4th IEEE International Conference on Computer Science and Network Technology, pp. 584- 589.
[19] Rajamohana S.P., Umamaheswari K., Dharani M., Vedackshya R., (2017), "A Survey on online review spam detection techniques", proceedings of IEEE International Conference on Innovations in Green Energy and Healthcare Technologies, pp. 8- 13.
[20] Liu P., Xu Z., Ai J. , Wang F., (2017), **"**Identifying Indicators of Fake Reviews Based on Spammer's Behavior Features**",** proceedings of IEEE International Conference on Software Quality, Reliability and Security, pp. 396- 403.
[21] Chauhan S.K., Goel A., Goel P., Chauhan A. and Gurve M.K., (2017), "Research on Product Review Analysis and Spam Review Detection", proceedings of IEEE 4th International Conference on Signal Processing and Integrated Networks (SPIN), pp. 399- 393.
[22] Christopher S.L. and Rahulnath H. A., (2016), "Review authenticity verification using supervised learning and reviewer personality traits", proceedings of IEEE International Conference on Emerging Technological Trends, pp. 16- 23.
[23] Shojaee S., Azman A., Murad M., Sharef N. and Sulaiman N., (2017), "A Framework for Fake Review Annotation", proceedings of 17th IEEE Computer Society UKSIM-AMSS International Conference on Modelling and Simulation, pp. 153- 159.
[24] Ahsan M.N.I., Nahian T., Kafi A.A., Hossain I., Shah F.M., (2017), "An Ensemble approach to detect Review Spam using hybrid Machine Learning Technique", IEEE 19th International Conference on Computer and Information Technology, pp. 381- 388.
[25] Ohana B and Tierney B, "Sentiment classification of reviews using SentiWordNet" , 9th. IT & T Conference. 2009: pp 1232-1243

Author Profile

Pankaj Chaudhary has completed his B.Tech and M.Tech, he is Ph.D(CSE) reasearch scholer at ICFAI University, Dehradun . He has published 16 National and International research papers in jounrals of repute. He has also attended several conferences. Currently he is doing his reasearch in analysing the genuinity of the online reviews of hotels.

Dr. Anurag Aeron is Associate professor(CSE) at ICFAI University, Dehradun, He has completed his Ph.D from IIT Roorkee. His research areas are Remote Sensing and GIS, Open Source Systems, Disaster Management, AI, Android Operating System, Machine Learning, IOT, NLP.

Dr. Sandeep Vijay, Working as Director at Shivalik College of Engineering, Dehradun, He has completed his Ph.D from IIT Roorkee. He has Proficiency in spearheading overall strategic research and developments projects, right from planning, cost controls,, resource mobilization, structured communications to final reviews, within cost & time parameters.