

An Algorithm for Deciding the Online Reputation of Hotels

Pankaj Chaudhary, Anurag Aeron, Sandeep Vijay

Abstract: After revolution in cell phone industry expansion and offering of promotional data packs by telecom companies like Reliance Jio, Airtel, Idea, Spice etc accessibility to the Internet has become very easy for the people. maximum people are now connected through social media viz. facebook, twitter, instagram etc. People are sharing their best and worst experiences for any brand. Various online review sites like Treebo, Yelp, Google Maps, and Tripadvisor OYO, Makemytrip, goibibo etc are used as an important source for the success of hotel businesses. Word of mouth has always been a powerful tool for marketing a business, Online reviews are today's word of mouth marketing, and these can make or break your business; In this research paper it is proposed for analyzing online reviews about hotels our algorithm must able to detect and analyzing fake reviewers based on user, tweet, timestamp, IP, collision and manipulation concept as well as to develop optimal model (based on group theory) for detecting fake reviewers, Improvement in enhancing sentimental analysis and the review detection model which can be implemented on all positive or all negative reviews, also the algorithm must able to identify the best fit of four machine learning techniques: (supervised machine technique, text mining technique, support vector machine learning technique and Naïve bayes machine learning technique) for specify and verify the different parameters of classification of reviews. Algorithm must able to Quantify the results of above techniques and extract the parameters to analyze the Genuinity of reviews based on Location, Security, Price, Quality, Ambiance etc.

Keywords: Hype, Quantification, collision, manipulation, machine learning, mining, deep learning etc

I. INTRODUCTION

1. Importance of Online Reviews: Online reviews are important due to various reasons:

Accessibility- Online reviews are incredibly easy to find, due to search engines, mobile apps, social media websites like facebook, twitter, LinkedIn etc.

Visibility- As a latest trend, online reviews for a business become visible to consumers before they even click a link. The ratings for each business are included in the results.

Online Reviews Help Your Business- Making positive information about any business available online can help in marketing efforts.

Search Engine Optimization- Popular online review sites like Facebook, Makemytrip, Yelp and Trip Advisor and Search engines like Google give favor to sites to rank higher in search results than its actual website.

Revised Manuscript Received on July 07, 2019.

Pankaj Chaudhary, Research Scholar, FST, ICFAI University, Dehradun, India.

Dr. Anurag Aeron, Associate Professor, FST, ICFAI University, Dehradun, India.

Dr. Sandeep Vijay, Director, Shivalik College of Engineering, Dehradun, India

Buyer Decision-Making- Maximum consumers use online reviews as important as recommendation given to them in person.

II. VARIOUS WAYS TO GET MORE ONLINE REVIEWS

Following techniques must be incorporated to increase the reviews about the products or business

- (1) Ask for Reviews
- (2) Make it Easy to Get Feedback
- (3) Engage Customers Online
- (4) Respond to Negative Reviews
- (5) Review Other Businesses
- (6) Be Forward-Thinking
- (7) Respond to Negative Social Media Engagement Quickly
- (8) Facilitate One-on-One Treatment
- (9) Don't Let Reviews Scare You
- (10) Be Kind to Yourself
- (11) Preparing to Respond to Negative Reviews
- (12) Find the root of the problem.
- (13) Get More Positive Reviews

III. LITERATURE REVIEW

Visani and Jadeja [1] has proposed the concept of a feeling classifier with the help of four machine learning techniques viz. supervised machine learning technique, text mining machine learning technique, support vector machine learning technique and Naïve bayes machine learning technique. This technique analyzes the previous reviews and can decide positive, negative and nonpartisan assumptions for a particular record. However the limitation with this method is that it does not fully utilize multiple evaluations of surveys. This method also can not conclude other outcomes except classifying a review as positive, negative and nonpartisan. The criterion and techniques for accommodating criticisms are also limited and static. Time stamping is also not considered.

Xue, Li, Seo and Pluretti [2] have focused on fake reviewers, according to them fake reviews are mostly generated by the fake reviewers. They concluded that Fake reviewers are never much trustworthy. They gave a trust based rating prediction system of reviewers in which they used proximity as trust weight; It was resolved that social relationships and friend lists can be used for review rating importance and prediction.



An Algorithm for Deciding the Online Reputation of Hotels

A trust-aware detection model was also proposed by them based on rating variance that iteratively calculates user-specific overall trustworthiness. However decision parameters for identifying the fake reviewer is highly static and based on pre-assumptions. Fake reviewer behavior may change with time; hence additional factors must be included especially average reviews generated, different segments reviewed by the reviewer etc.

Jiang and Cui [3] concluded that unusual behavior is an indication of spam and proposed a behaviour detection techniques based on the gathered information from the data. According to them facte(user, tweet and IP), size and density play important role at twitter to detect the situation when fraudsters try to manipulate the most popular tweets. Concept is good but it can be successful only when numbers of reviews are very high. quantifying the suspiciousness of a behavioural pattern is still an open issue. Analysis of past trustworthiness of the user must also be done.

Rout, Dalmia and Choo [4] discussed 4 important approaches of semi supervised machine learning method that can be used to improve the F-score metric in classification. In addition to the traditional approaches some new dimensions were incorporated (1) Parts-of-Speech (2) Linguistic (3)Word Count Features (4)Sentimental Content features.

This model can give achieve the F-score of 0.837. However further improvements are possible by implementing and evaluating by collecting the data from various websites in real-time. In addition to the textual content, multimedia content must also be considered.

Deng and Chen [5] consider the reviews in four dimensions: taste, service, environment and overall attitude. result of the four dimensions is analyzed and found consistent, then review will be of hype review category. The major problem in this model is that more than 60% of errors occur mainly due to two reasons. (1) Incomplete sentence-breaking is possible by using of comma as a delimiter. (2) Few neutral reviews can be wrongly sorted by several sets of training, and therefore it becomes very difficult to judge whether they are hype or not.

Ruchansky, Seo and Liu [6] presented CSI model composed of : Capture, Score, and Integrate, Capture is based on the response and text; with the help of Recurrent Neural Network in a given article it capture the temporal pattern of user activity. Score module learns the basic characteristic based on the users behaviour, and these two modules are finally integrated with the third module to analyse the genuinity of an article. This CSI Model needs further improvements with the help of reinforcement learning and crowd sourcing.

Mukherjee, Venkataraman, Liu, Glance [7], utilized Amazon Mechanical Turk (AMT) to in-depth investigation of supervised learning for fake review detection. However this model was failed at Yelp. To improve classification on Yelp's real-life data, a set of behavioral features were also to

be considered which resulted in a major accuracy improvement.

Kokate, Tidke [8] proposes a model based on behavioural approach to detect review spammers. They derive an aggregated behaviour scoring methods for ranking of the reviewers

However exploring different ways to learn behaviour patterns to improve the accuracy of the current regression model is also needed.

Kolhe, Joshi, Jadhav and Abhang [9] proposed a model that is based on the concept of detecting fake reviewer groups. Individual fake reviewer detection is a tough task however fake groups detection is an easier task. This model proposed the scoring algorithm based on the process of frequent item set mining method.

Kale, Jadhav and Pawar [10] proposed a fake review detection method based on the comments as the indicators like a discontinuous flow of text, inadequate and in-appropriate language or the content not related to the specific context and check similarity and difference among the comments.

This system is purely based on natural language processing techniques and hence this method is highly static and pre-assumption based, it cannot consider different age group reviewers or different mindset users.

Adike and Reddy [11] proposes a behavioural approach based method to identify review spammers. This method mainly focuses on content review centric spam identification techniques. Different ways to learn can improve this model as well as we may also quantify behaviour patterns related to the spamming. It will improve the accuracy of the existing regression methods.

Bonde, Kharabi and Sabale [12] proposed opinion integration algorithm. This algorithm includes following spam detection methods
-opinion spam detection,
-opinion summarization,
-opinion visualization
-opinion assessment.

However this method can be implemented only when quantity and varieties of the reviews are available in bulk.. This classifier need to be more robust and consider all situations.

Crawford, Khoshgoftaar, Prusa, Richter and Najadain [13] proposed comparison of various machine learning techniques;

However, obtaining labelled reviews is difficult. Special consideration need to be done to resolve how many actual features are required



and which features are most beneficial.

Elmurngi and Gherbi [14] have applied Text classification and sentimental analysis methods on a real world dataset of movie reviews with the help of four supervised machine learning algorithms and Decision Tree (DT-J48) without stopwords and with stop words. This work has used only specific kind of data sets. Different feature selection methods are also not utilized. Furthermore, collision and manipulation issues are not addressed properly.

Fontanarava, Pasi and Viviani [15] proposed a supervised classifier based model that utilized random forests (RF) concept by considering both old as well as new features. A large-scale labeled YelpNYC dataset is used to extract these features. The strengths of YelpNYC dataset are:

- The quantity of reviews generated per user,
- The diversified kinds of entities reviewed,
- The datasets contain following general information,
- the content,
- label,
- rating,
- date of each review when it was generated.

This allows to generalize the proposed analysis to different review sites. This model can be further improved by proposing some new techniques to collect more and more public reviews as it is successful only on bulk of reviews. Variety of reviewers is also not considered much. However this model assumes that dataset used in evaluations contains positive as well as negative opinions. Scope for integrating of minimal meta-data for reduction also existed.

Wahyuni and Djunaidy [16] has proposed to detect fake reviews based on the proposed system termed as (ICF++) that measures three things

- The honesty value of reviews,
- The trustiness index of the reviewers
- The reliability index of a product.

The honesty index of reviews is measured by implementing the opinion mining and text mining techniques.

However the use of only ratings to assess the review fakeness or genuine is inadequate,

Lin, Zhu, WuI, Zhangl, Wang and Zhou [17] proposed a model that is based on identification of fake reviews with the help of review sequence, that consists of the ordered reviews according to their posted time points. This model is based on three steps:-

1. With the help of six time sensitive features highlight the fakeness of review as per the review contents as well as the reviewer behavior.
2. Devised the supervised solutions and as well as threshold-based for spotting fake review.
3. Conduct a number of experiments on a set of actual review to analyze six proposed features.

By increasing or decreasing the sample size and iterations as per the situation and importance this model can further be improved.

Li, Feng, Zhang, Li [18] have proposed semantic and emotion model that includes consideration for review

density, emotion mathematical modeling and semantic techniques for fake review detection.

However for improvement in the model we may collect abundant data based on review from several review websites, computer based labeling of different reviews may also be incorporated to reduce the workload.

Yin, Wang and Liu [19] proposed model based on Opinion mining, emotional analysis, subjective comments recognition, and knowledge discovery.

This method consists of four parts viz,

- opinion spam detection,
- opinion summarization,
- opinion visualization and
- opinion assessment.

However improved classification methods may be incorporated in this model to reduce the overhead incurred.

Rajamohana, Umamaheswari, Dharani, Vedackshya [20] divided reviews into three types: Type I, II and III reviews. Using various machine learning techniques a detailed survey is done with a purpose for detection of spam and genuine reviews. However this model is good only for the situations when reviews are provided by different mindset people. This model can be further improved by doing research and improvement in big data approaches to reduce computational complexity and the set of features.

Liu, Xu, Ai and Wang [21] proposed that fake reviewer has common type behavioural features, based on these features; eight identification indicators were identified to detect spam reviews of the product.

They present algorithms to recognize similar reviews and relevant reviews, respectively. However further improvements in indicators are possible to recognize more non-relevant reviews.

Chauhan, Goel, Chauhan and Gurve [22] proposed a model to incorporate sentiment analysis method for fake reviews detection. In this method first of all a big own dictionary having sentiment words is prepared. Words are assigned the weight age also.

Then using natural language processing a method is proposed based on shallow dependency parser for calculation of the sentiment score of the reviews. Various discriminative rules are proposed through past observations. These rules are added with the various time series methods to find out the spam and fakeness of the reviews. Continuous updation of the dictionary containing sentiment word and associated rules flexibility as per the situations is a big challenge.

Christopher and Rahulnath [23] proposed a model for fake review analysis based on personality prediction. The fake reviewer will try significantly to deviate from the reviews given by usual public opinion and tactfully try to influence the customer to make opposite choices with a very convincing style.

An Algorithm for Deciding the Online Reputation of Hotels

The Big 5 model is used in this model. This model can be further improved with the help of linking of the different reviewer ids in product reviews.

Fei, Mukherjee, Liu¹, Hsu, Castellanos and Ghosh R. [24] proposed a model based on nature of reviewers in a burst. It classifies reviews based on a different set of features. But specification of good and bad features need novelty in the approach.

.Shojaee, Azman , Murad, Sharef and Sulaiman [25] proposed a framework to annotate review corpora for fake review detection. They selected most effective features using feature selection – wrapper method to develop two set of clues, in form of questions, for spam and spammer detection. An algorithm is developed to calculate label for each review based on provided hints for each reviewer and his reviews. Further improvement is possible in the proposed framework by investigating more hints and evaluating our system using bigger dataset. Also, giving weight to each feature to highlight the role of the most effective features on the labeling process to decrease misclassification.

Ahsan, Nahian, Kafi, Hossain and Shah [26] presented a model for the purpose of detecting the review spam by combining 2 Active & Supervised learning for both real life as well as fabricated data. Further improvements are possible that may include experiments on large-scale datasets received from various domains of several languages.

Ahsan, Nahian, Kafi, Hossain and Shah[27] proposed active learning approach based on the TF-IDF features that uses classification Linear SVM, SGD and Perceptron. This model may further be improved for adding large-scale datasets received from different types of domains to increase the size and diversity and variety of the data. It may also be improved for several diverse sets of tuning and smoothness techniques for n-gram models

IV. OPINION SURVEY

A questionnaire was distributed among 358 persons who have taken decision in past based on the reviews to know their opinion about the factors that they keep in mind while deciding online reputation of any brand. Following questions were asked in the survey:

Q1: Which of the following are the major criteria to decide the trust on review? Some of these are:

SN	Parameters	First Choice by
A	Past Profile of reviewer who give recommendation.	89/358[25%]
B	Similar Rating by the reviewers of different age groups and from different regions.	38/358[11%]
C	Difference of the time duration	31/358[9%]

	among the reviews.	
D	Grammar and gravity of the statements.	41/358[12%]
E	Syntactical analysis of the statements.	56/358[16%]
F	Categorizing the best, average and worst reviews and conclusion	103/358[29%]



Fig 1: First factor that affect the online trust

Q2: Do they consider following parameters to decide the trust on review?

SN	Parameters	Do Consider
A	Past Profile of reviewer who give recommendation.	141/358[40%]
B	Similar Rating by the reviewers of different age groups and from different regions.	96/358[27%]
C	Difference of the time duration among the reviews.	58/358[16%]
D	Grammar and gravity of the statements.	106/358[30%]
E	Syntactical analysis of the statements.	112/358[32%]
F	Categorizing the best, average and worst reviews and conclusion	296/358[83%]

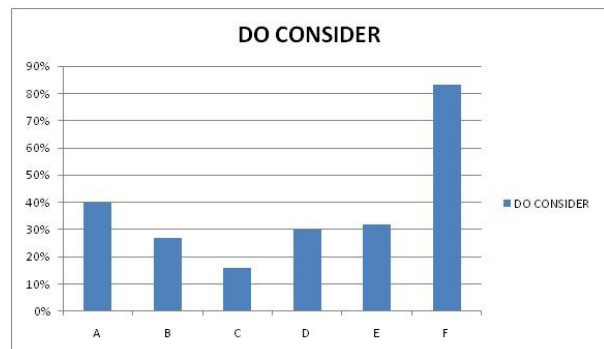
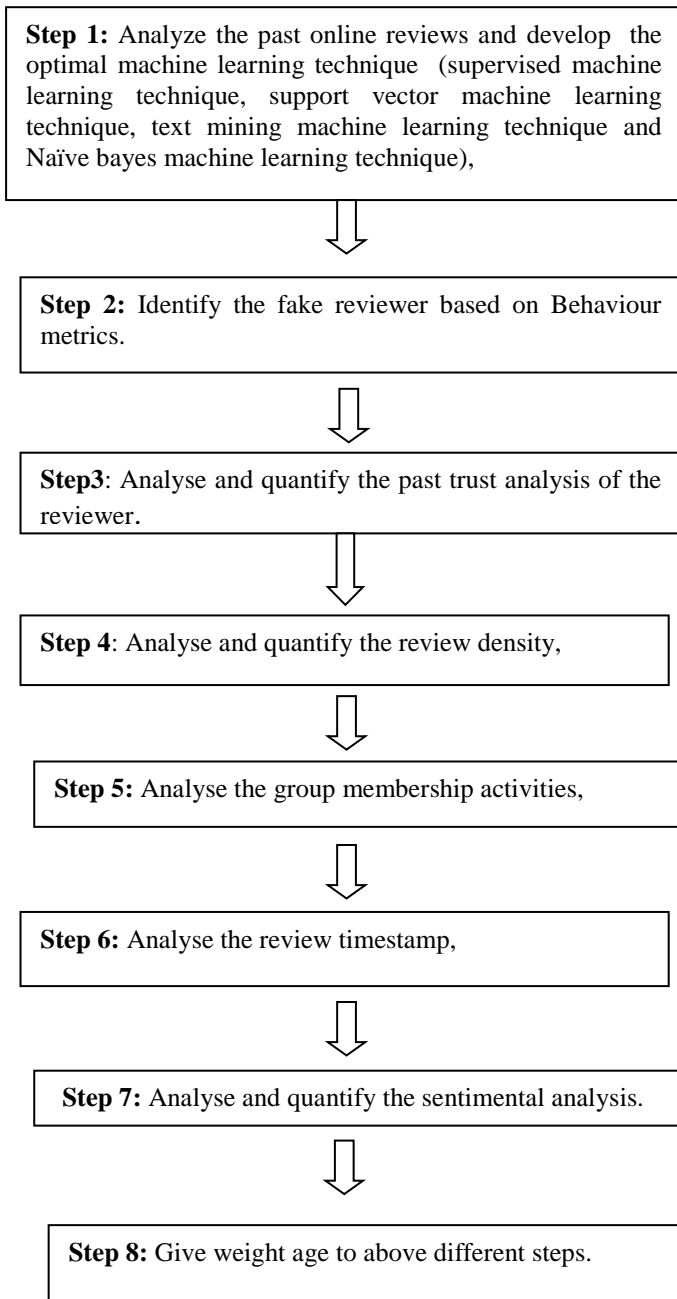


Fig 2: Factors that affect the online trust



V. 5. PROPOSED MODEL

Based on the above literature survey and conducted survey a novel algorithm is proposed that is highly flexible and implementable dynamically in all the situations as per below steps:



Based on above 8 steps conclude the appropriate parameters to analyze the genuinity of hotels based on Location, Security, Price, Quality, Ambiance etc.

VI. CONCLUSION AND FUTURE WORK

The scope is to develop novel model to help the users to make analysis of past online reviews and to conclude the various parameters about any hotel viz. Location, Security, Price, Quality, Ambiance. This will collect several past online reviews about the hotels from multiple websites like Tripadvisor, OYO, Treebo, Yelp, Crewel, Makemytrip, goibibo etc. Although this work is specifically proposed for

helping customers in selection of the best hotels by analyzing the previous online reviews, Yet the similar model may be designed after minor modifications for taking right decision in selecting the best colleges, best products etc.

REFERENCES

1. Visani C. and Jadeja N., (2017), "A Study on Different machine learning techniques for spam eview detection", IEEE Transaction 978-1-5386, pp. 1887-1892.
2. Xue H., Li F., Seo H. and Pluretti R., (2015), "Trustaware 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., Dalmia A. and Choo 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. 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. Kolhe N.M., Joshi M.M., Jadhav A.B., Abhang P.D., (2014), " Fake Reviewer Groups' Detection System", IOSR Journal of Computer Engineering (IOSR-JCE). e-ISSN: 2278-0661, p- ISSN: 2278-8727 Volume 16, Issue 1, Ver. V (Jan. 2014), pp. 06-09.
10. 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.
11. 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.
12. 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.
13. 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.
14. Elmurngi E. and Gherbi A.,(2017), "An Empirical Study on Detecting Fake Reviews", proceeding of IEEE The 7th International Conference on Innovative Computing Technology, pp. 107-114.
15. 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.
16. Wahyuni E. D. And Djunidy 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.
17. 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.

An Algorithm for Deciding the Online Reputation of Hotels

18. 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.
19. 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.
20. 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.
21. 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.
22. 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.
23. 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.
24. Fei G., Mukherjee A., Liu B., Hsu M., Castellanos M., Ghosh R. "Exploiting burstiness in reviews for review spammer detection", Proceedings of the Seventh International AAAI Conference on weblogs and social Media, (2013), pp. 175- 185.
25. Shojae S., Azman A., Murad M., Sharaf N. and Sulaiman N., (2017), "A Framework for Fake Review Annotation", proceedings of 17th IEEE computer society conference on Modelling and Simulation, pp. 153- 159.
26. 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 Nineteenth International conf. on computer and Information Technology, pp. 381- 388.
27. Ahsan M.N.I., Nahian T., Kafi A.A., Hossain I., Shah F.M., (2016), "Review Spam Detection using Active Learning", IEEE 19th International conf. on computer and Info Technology, pp. 368- 375.

AUTHORS PROFILE



Pankaj Chaudhary has completed his B.Tech and M.Tech, he is Ph.D(CSE) research scholar at ICFAI University, Dehradun . He has published 16 National and International research papers in journals of repute. He has also attended several conferences. Currently he is doing his research 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, risk mitigation, resource mobilization, structured communications to final reviews, within cost & time parameters.