Incentivized Comment Detection with Sentiment Analysis on Online Hotel Reviews

Md. Niaz Imtiaz[1], Md Toukir Ahmed[1,1], and Antara Paul[1]

[1]Pabna University of Science and Technology

July 24, 2020

Abstract

With the enormous platforms available in present days, consumers communicate and interconnect online with web users all around the world to share their experiences. Thus, online platform has become a major source of reviews about different entities. People presently travel frequently around the world for different purposes. Seeking good hotels for accommodation is a prime concern. Customer reviews on hotels help future customers to take decisions about their accommodation as well as help hotel owners to rethink about designing customer facilities. However, many online reviews are biased due to different factors. Many hotel owners come up with attractions like referral rewards, coupons, bonus points etc. to the reviewers to motivate them in writing biased reviews. We have worked on US's 100 hotel and found 952 incentivized reviews out of 19175 reviews, which is 4.96% of total reviews. A categorization on incentivized reviews is performed as well. Furthermore, hotels are distinguished based on real and incentivized reviews found on them. Results are verified using machine learning algorithms. Random Forest, K-Nearest Neighbor and Support Vector Machine are applied as machine learning algorithms to validate the accuracy of our model and their prediction results are compared. Random Forest outperforms with 94.4% prediction accuracy.

1. Introduction

Now a days the acceptance and significance of social media sites has been raised in people's day to day life. People share their feelings via social media through posting reviews [9]. As a result, a wide range of knowledge is available to customers while making decision for booking of a hotel, including electronic post and reviews given by consumers. Because of this fluency of information, consumers gain adequate knowledge by reading the other consumers' satisfaction or regret with a hotel booking experience from online [8].Analysing the large volume of online reviews available would produce useful actionable knowledge that could be of economic values to vendors and other interested parties [6]. Consumers can check the star rating as well as read the other customers' experiences. There are lot of ways by which consumers can communicate like blogs, social media forum discussions etc. However, more and more incentivized comments are also being posted in user's account on the social media which are usually biased by companies, retailers or owner of any hotel offering some advantages for comments made by customers and electronic post in social media. In research it is found that online review websites are mostly affected by frequent communication effort [5]. Researchers have found instead of company provided information consumers rely more on information given by consumers. Thus, after family and friends' online reviews have become the second most trusted source of information [14]. As a result, often consumers do not get their expected services and they become disappointed with theincentivized reviews. While posting fake consumer reviews is hard to control, this situation has resulted in manufacturers, retailers and third-party companies developing technology platforms for systematically managing customer's real and incentivized reviews [20]. Research on the effectiveness of these marketing campaigns is scarce. There are few studies that offer an explanation on how consumers respond to these techniques or if they might perceive any of them as disingenuous [6, 20]. The objective of this study is to differentiate proper reviews and incentivized reviews. This study might help people who make their decisions on online reviews.

2. Preliminaries

2.1. Natural Language Processing

The main aim of Natural Language Processing is to understand human level languages. Methods of NLP for processing are perceptive and effective. Indicative software engineering processes that provide grounds for applying natural language processing. The increasing adoption of outcomes from the Natural Language Processing (NLP) community in software engineering research comes in line with recent advances in NLP which provide opportunities for efficiently processing large amounts of data. Examples of research topics include the identification of textual characteristics reflected in review comments, the assessment of the quality of any products, the detection and classification of different types of emotions [12]. Most importantly, though: the subject of NLP— language—is a proxy for human behavior, and a strong signal of individual characteristics. People use this signal consciously, to portray themselves in a certain way, but can also be identified as members of specific groups by their use of subconscious traits [7]. NLP often relies on statistical techniques, specially to formulate the words in texts. With the help of NLP techniques, rule-based methods are now being experimented in different new approaches that hold resource like knowledge. The scope of NLP is great and every day the number of tasks performed by NLP is increasing. NLP applications includes semantic analysis, question answering, chatbots, automatic summarization, market intelligence, opinion mining, language translation etc. Opinion mining as a research discipline with NLP has emerged during last 15 years and provides a methodology to computationally process the unstructured data mainly to extract opinions and identify their sentiments. [16]. Firstly, using formal grammar and lexicon, text is parsed syntactically. Then, NLP techniques are applied to interpret semantically to understand what the text is actually saying. NLP comprises of techniques like tokenization, part-of-speech tagging, word stemming, lemmatization, multiword phrase grouping, synonym normalization, word-sense disambiguation, and anaphora resolution and role determination.

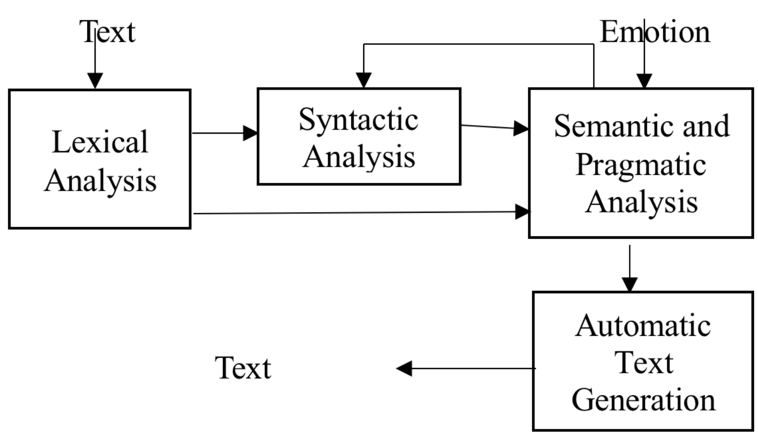

Figure 1. Steps of Processing Natural Language

Figure 1 shows the natural language processing steps. At lexical analysis step, meaning of individual words are interpreted. Then a series of processing steps are followed to understand word-level. Each word is assigned by corresponding POS tag. In syntactic phase, words are analyzed to understand the sentence and its grammatical structure. A parser along with a grammar is needed for syntactic analysis. The output of syntactic analysis shows the inter-relationship between words with respect to structure.

2.2. Sentiment Analysis

Sentiment analysis measures polarity of sentiment for a given text or opinion. It also aims to detect the subjectivity of text as well. Sentiment analysis has been applied to various software engineering (SE) tasks, such as evaluating app reviews or analyzing developers' emotions in commit messages [3]. In this work, subjective opinions on specific products made by customers are recognized using sentiment analysis to understand customer perception. Sentiment analysis identifies human attitude towards products or services by recognizing the polarity of opinion that he or she has given. Today text classification or review analysis research starts from designing the best feature extractors to choosing the best possible machine learning classifiers [1]. Unstructured text data produced on the internet grows rapidly, and sentiment analysis for texts becomes a challenge because of the limit of the contextual information they usually contain. Figure 2 shows different analysis methods for sentiment scoring.

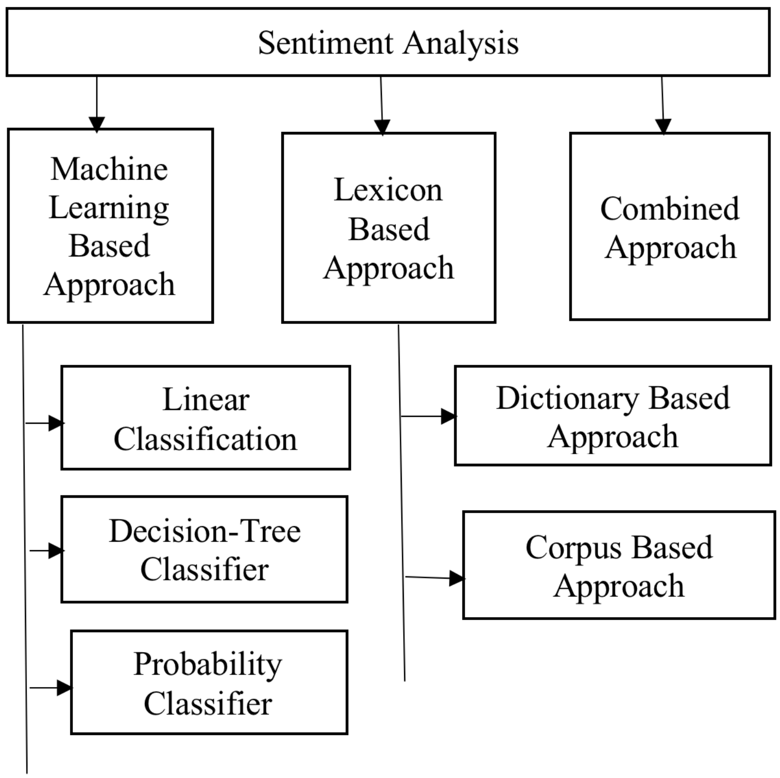

Figure 2. Sentiment Analysis Approaches

Lexicon based approach measures the sentiment of text or opinion and classifies the polarity into positive, negative or neutral utilizing sentiment lexicon. This approach is more understandable and can be easily implemented. It is easy and more understandable to implement Lexicon based approach. There are two classification of lexicon-based approach: corpus-based approach and dictionary-based approach. The Twitter messaging service has become a platform for customers and news consumers to express sentiments that include dictionary-based approach and use supervised machine learning tools for sentiment classification [11]. This research follows the supervised machine learning approach.

In our work we have used dictionary-based approach for lexicon analysis. In machine learning based algorithms the quality and volume of training data affect the prediction performance of the classifier. A large database is required to get good result. In our work we have used machine learning for validating our work.

A several systems have been experimented in detecting spam such as The Naive Bayes classifier and tfidf (term frequency – inverse document frequency). But the drawback of those systems is that they do not extract semantic information of the phrases and spam words. But analyzing text of public comments has the significance or retrieving semantic information from the words. For that some sentiment analysis algorithms named Support Vector Mechanism (SVM), K-Nearest Neighbors (KNN), Vader Lexicon, Decision Tree, and Random Forest have been checked. Comparing all algorithms' work efficiency, advantages and disadvantages, Vader algorithm is found appropriate for finding sentiment score of this research because it performs better than most of the other tools. The algorithm considers several important factors like capitalization, excess of punctuation that lead to a better performance compared to other approaches those ignore them. It also improves the accuracy of the sentiment polarity score for a review.

Table 1 .Tools for Analysing Sentiment

Tools	Work
Feelings Polarity	Based on the reviews or emoticons.
Linguistic Inquiry and Word Count	By using dictionary, it recognizes emotional, cognitive, and structural components of t
Happiness Index	It gives the happiness index of the analysed text between 1 and 9 using Affective Nor
SentiStrength	It gives polarity score to phrases in the context (how much positive or negative) using
SentiWordNet	It recognizes the sentiment using WordNet where nouns, adjectives, verbs are assigned
VADER	VADER classify sentences using sentiment lexicon. All the necessary words are stored

There are several tools for analysing sentiment [Table 1]. For measuring sentiment polarity, we have used VADER, since it is efficient and simple.

2.3. Text Mining

The majority (80% approx.) of today's online data are unstructured. So, it is often difficult for computer system to process such unstructured text. To extract meaningful information from unstructured text some useful techniques are needed. Meaningful text derived by these techniques are then stored in text database. The source of text can be emails, SMS, chats, journal articles, newspapers, product or service reviews. Most of data stored in different organizations, institutions and industries are in electronic form. Text data mining can be viewed as knowledge discovery from unstructured text, that is extracting precious information from unorganized text [10]. Using different techniques unstructured data are interpreted into human or machine-readable form. Text mining concerns with text analysis, information extraction, clustering and visualization. That is why it is called a multidisciplinary field.

2.4. Incentivized Reviews

The rise of internet has enhanced people's way of communication more efficiently. E-commerce is one of the major inventions in marketing sector. Businesses now have developed globally in large scale with much less efforts compared to the strategies twenty or thirty years back. Consumers don't need to visit markets or shops physically, instead they can choose their products among a large collection product from home. Purchasing a desired product is just a matter of clicking mouse or pressing enter key of keyboard. Nevertheless, the fundamental concept of commerce has not been changed yet [17]. There are still a large number of buyers and sellers, product related issues and issues of being satisfied or disappointed with product are present. Presently, people often rely on online reviews while making a decision to purchase thinking that the reviews are unbiased and real. Customer reviews generate more sales, affect consumers trust and create more word-of-mouth spill over effect [6][5]. But the shocking thing is that most often the customers get disappointed for the reviews that are not real. Fake or biased reviews spreading is a great threat for consumers [18]. Different government and private agencies are trying to protect customers from being deceived. These economic hazards are threatening the billions in e-commerce revenue. A website, reviewskeptic.com, was developed by Cornell University researchers to identify fake hotel reviews with approximately 90% accuracy. Some businesses are even trying to remove their name from review websites or take action against the trend of fake reviews, by encouraging their customers to post funny, unfavourable reviews for them [17]. Hence filtering out the reviews has become a prime need. So, it is necessary to detect biased reviews from hotel because day by day the use of hotels is increasing.

2.5. Related Works

Researchers are working in last few years on text and online review processing to determine the sentiment polarity that help consumers with comparison of products and services. Very few researches have been conducted to detect the frauds in online reviews. This work aims to understand and analyse the characteristics of fake news especially in relation to sentiments, for the automatic detection of fake news and rumours [15]. Salehan and Kim analysed the forecasters of readership and usefulness of online customer reviews using a sentiment mining approaches [13]. They found titles consists of highly positive reviews have more readerships. Wei Wang, Hongwei Wang and Yuan Song proposed a model that enables the detection of aspect ranking from online reviews [19]. They measure the influence of user opinions by information gain theory rather than sentiment strength alone.

Kostyra et al. applied a choice-based conjoint experiment in analysing consumer reviews [4]. They combined all relevant levels of online consumer reviews and that identified the effect of online reviews on customer choice. They found that consumers' choices are not influenced by the volume and variance. They tried to moderate the impact of valence on consumer's choices. Costa et al. experimented a data mining technique to predict incentivized reviews based on some selected features such as the length of reviews, how helpful they are etc. [2].[14] also used a text-based approach for fake news detection but considered the test, response and clustering of user features determined by support vector decomposition and integrated into a hybrid model.

A number of works have been conducted on online customer reviews analysis and determining their impact on customer's choice. But incentivized or biased reviews can mislead customers. Determining the frauds that are frequently occurring in online platform is a new concept. In this research we analysed this perspective.

3. Materials and Methods

Dataset in this experiment is the collection of public reviews in text format on different hotels in United States. Public reviews on a list of 100 hotels at different towns in U.SA are extracted from renowned Kaggle website. Each hotel has more than 500 reviews. From the dataset we used only four columns and the rest of the columns have been removed for the simplicity of our work. Sample of our dataset is given in Table 2.

Table 2 . Dataset

Hotel Name	Review
Rancho Valencia Resort Spa	Our experience at Rancho Valencia was absolutely perfect from beginning to end!!!! We felt s
Days Inn and Suites Albany	In my line of work, I use meeting space in hotels often.
Hotel Phillips	Old hotel with many remaining architectural charms and most modern amenities. The staff i

In table 2, customers have expressed their feeling about the hotel they experienced via comment.

VADER algorithm is experimented to calculate sentiment polarity. uses subjectivity and polarity concept. Polarity concept defines positive and negative words as well as it defines the range of polarity which is in between -1 and 1. Polarity, also known as orientation is the emotion expressed in the sentence. Subjectivity tells when a text is an explanatory article which must be analysed in context. Previous experiment on VADER showed remarkable and precise results. In the field of public comment analysis where the text is complex with mixture of variety of text, VADER is found to perform well. Based on a sentiment lexicon and a grammar VADER analyses the sentiment polarity of a sentence or opinion. The words in the lexicon rated as negative or positive and how negative or positive as well based on public given score. To determine the positivity or negativity of words the developers of these approaches need to get a bunch of people to manually rate them. VADER generates four sentiment metrics from these word ratings. The first three represent the proportion of negative, neutral and positive of the word and the last one shows the compound score calculated from the sum of ratings. The compound score is normalized between -1 (most extreme negative) and +1 (most extreme positive). Sentiment is categorized as-

Positive sentiment (compound score >= 0.05)

Neutral sentiment (compound score in between {-0.05, 0.05})

Negative sentiment (compound score <= -0.05)

Table 3 . Sentiment Score

Compound	Positive	Negative	Neutral
0.431	0.192	0	0.808
0.848	0.199	0.098	0.703
- 0.296	0	0.355	0.645

Table 3 shows how VADER scored three random sentences. The scores in the columns positive, negative and neutral indicate how much a sentence is positive, negative and neutral respectively.

Algorithm
1: Import the dataset. 2: Select the required column discard the others. 3:**For each hotel:** Calculate the sentiment score c

After calculating sentiment compound score for each comment of a hotel, we calculated the *mean sentiment score* for each of the hotel. The*median sentiment score* is also calculated for each hotel.Then *standard deviation from mean* and *standard deviation from median* are calculated separately and compared with each other. Finally, sentiment score for each hotel is identified based on the deviation. If *standard deviation from mean* is less, *mean sentiment score* is selectedas final sentiment score otherwise*median sentiment score* . We always tried to find the optimal value so that the variance becomes low.

Table 4. Mean, Median and Standard Deviation Values

Mean	Median	STD from Mean	STD from Median
0.225	0.401	0.639	0.663
0.431	0.494	0.735	0.524

Table 4 shows the mean, median and standard deviation scores for two reviews of two hotels. The first one indicates that standard deviation from mean is less than standard deviation from median. So, in this case mean score should be selected to minimize variance.

With respect to the final sentiment score for a hotel variance for each of the reviews is calculated considering polarity score of that review and final sentiment score of the hotel. A threshold value of 0.6 is chosen. A review is marked as incentivized whose variance is more than the threshold value. This process continues for all the individual hotel reviews.

Finally, three well-known machine learning algorithms such as K Nearest Neighbor (KNN), Random Forest and Support Vector Machine (SVM) are applied to measure the accuracy. The entire dataset is divided into two sets. A set with 70% of total data is used for training and the set with rest 30% is used for test. The prediction results of the three algorithms are compared.

4. Result and Discussions

In this research natural language processing techniques are experimented to analyze sentiment. Since data is in unstructured format, instead of transforming it into numeric values, we recognized the semantic and contextual meaning. Python is used as a programming language. For mining text Natural Language Toolkit (NLTK) package is implemented and for manipulating data other packages like Pandas and NUMPY are included. We used VADER algorithm for sentiment score. We have used some statistical functions- mean, median and standard deviation for our work with respect to sentiment score.

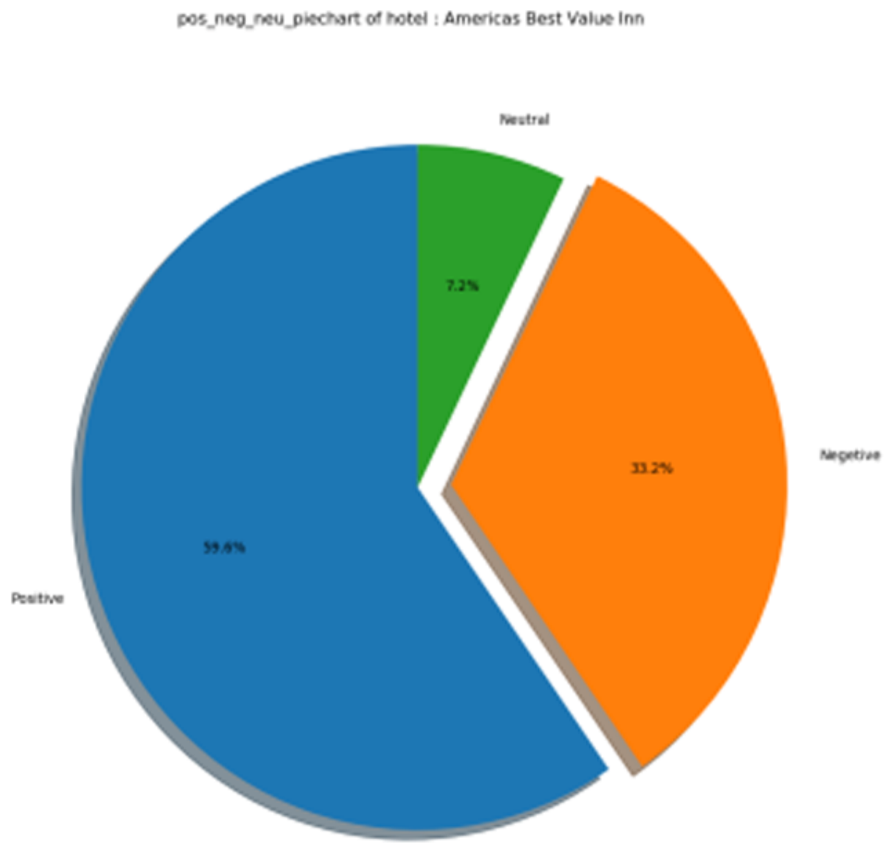

Figure 3. Polarity of Reviews on Americas Best Value Inn hotel

Polarity of reviews are measured for individual hotels. For example, Americas Best Value Inn hotel has 59.6% positive reviews and 33.2% negative reviews and 7.2% neutral reviews [Figure 3].

Name	Reviews	Compound	Mean_Compound	Med_Compound	STD_Mean	STD_Med	Small_Dev	Difference	Output
40 Berkeley Hostel	If you just m	0.431	0.2852	0.3561	0.103	0.053	0.053	0.0749	Real
40 Berkeley Hostel	The staff he	0.9682	0.2852	0.3561	0.483	0.4328	0.4328	0.6121	Incentivized
40 Berkeley Hostel	Avoid Paula	-0.096	0.2852	0.3561	0.27	0.3197	0.27	0.3812	Real
40 Berkeley Hostel	As an Ameri	0.4735	0.2852	0.3561	0.133	0.083	0.083	0.1174	Real
40 Berkeley Hostel	my room ma	0.7127	0.2852	0.3561	0.302	0.2522	0.2522	0.3566	Real
40 Berkeley Hostel	I love this pl	0.3337	0.2852	0.3561	0.034	0.0158	0.0158	0.0224	Real
40 Berkeley Hostel	We stayed f	0.746	0.2852	0.3561	0.326	0.2757	0.2757	0.3899	Real
40 Berkeley Hostel	Don't let the	0.5312	0.2852	0.3561	0.174	0.1238	0.1238	0.1751	Real
40 Berkeley Hostel	I stayed in B	0.1491	0.2852	0.3561	0.096	0.1464	0.096	0.1361	Real
40 Berkeley Hostel	walking dist	0.1154	0.2852	0.3561	0.12	0.1702	0.12	0.2407	Real
40 Berkeley Hostel	The cost of t	0.7501	0.2852	0.3561	0.329	0.2786	0.2786	0.394	Real
40 Berkeley Hostel	The location	0.4545	0.2852	0.3561	0.12	0.0696	0.0696	0.0984	Real
40 Berkeley Hostel	Considering	0.4309	0.2852	0.3561	0.103	0.0529	0.0529	0.0748	Real
40 Berkeley Hostel	We stayed h	0.7263	0.2852	0.3561	0.312	0.2618	0.2618	0.3702	Real
40 Berkeley Hostel	Alright. This	0.8716	0.2852	0.3561	0.415	0.3645	0.3645	0.5155	Real
40 Berkeley Hostel	40 Berkeley	0.7611	0.2852	0.3561	0.337	0.2864	0.2864	0.405	Real
40 Berkeley Hostel	I'd never sta	0.7967	0.2852	0.3561	0.362	0.3116	0.3116	0.4406	Real
40 Berkeley Hostel	Overall a ve	0.3574	0.2852	0.3561	0.051	0.0009	0.0009	0.0013	Real
40 Berkeley Hostel	Could have	0.3562	0.2852	0.3561	0.05	0.0001	0.0001	0.0001	Real
40 Berkeley Hostel	This is a che	0.7184	0.2852	0.3561	0.306	0.2562	0.2562	0.3623	Real
40 Berkeley Hostel	A nice place	0.7371	0.2852	0.3561	0.32	0.2694	0.2694	0.381	Real
40 Berkeley Hostel	I would not	-0.2523	0.2852	0.3561	0.38	0.4302	0.38	0.5375	Real
40 Berkeley Hostel	I have staye	0.7184	0.2852	0.3561	0.306	0.2562	0.2562	0.3623	Real

Figure 4. Output of Reviews

Figure 4 shows the output result for some of the reviews of 40 Berkeley Hostel. The columns are described below:

Hotel: Name of the hotel.

Review: Review on the hotel.

Compound: Compound polarity each review.

Mean_compound: The calculated mean compound score of all the reviews for each hotel.

Med_compound: The calculated median compound score of all the reviews for each hotel.

STD_mean: Standard deviation of *Compound* from

Mean_compound.

STD_med: Standard deviation of *Compound* from

Med_compound.

Small_Dev: Smaller value between *STD_mean* and

STD_med.

Difference: Unsigned difference between *Compoundand Mean_compound* or

Med_compound based on *Small_Dev.*

Output: Status of the review.

For each review there is an entry in the table. All the numeric values are rounded in three decimal points. From the first review on 40 Berkeley Hostel it is seen that polarity of that review is 0.431 approx. So, it is a positive review but not too much positive. Here standard deviation from mean (0.103 approx.) is less with compared to that of median (0.053 approx.). As standard deviation from median is less with compared to that of mean, the final sentiment score for that hotel will be median compound score. So, the difference is calculated between polarity of that review (0.431 approx.) and median compound score (0.3561 approx.). The review is real as difference is less than the threshold value (0.6 approx.). The second review is incentivized as difference (0.6121 approx.) is greater than threshold value.

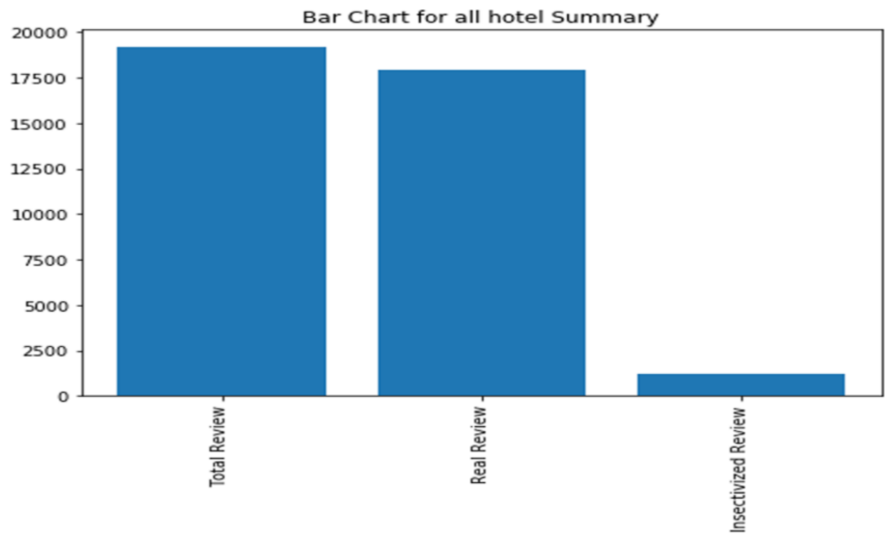

Figure 5. Summary of Reviews

Figure 5 shows the number of total number of reviews, real reviews and incentivized reviews. Out of 19175 total reviews we found 952 incentivized reviews and18058 real reviews. That is 4.96% of total reviews are marked as incentivized which is significant.

Experimenting K Nearest Neighbor, Random Forest, Support Vector Machine, the accuracy of our identification is measured. Our identified output is used as class level. So, class level has two possible values- real or incentivized.

Table 5. Accuracy Given by Algorithms

Algorithm	Accuracy
KNN	92.8%
SVM	93.2%
Random Forest	94.4%

The prediction accuracy on test data is shown in Table 5. Best accuracy (94.4%) is given by Random Forest algorithm while other algorithms have good accuracy levels too.

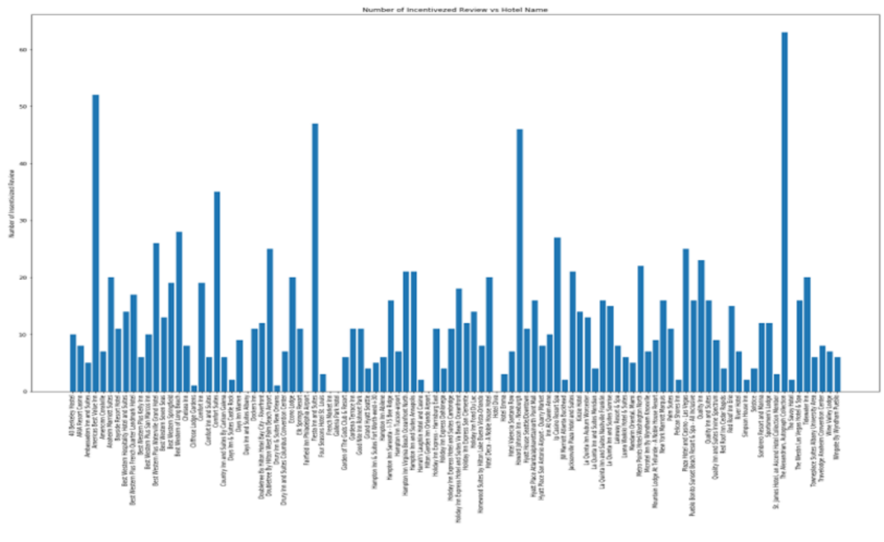

Figure 6. Incentivized Reviews for Individual Hotels

While looking at result closely number of incentivized reviews for individual hotels is monitored. It is found that some hotels have high number of incentivized reviews while some have no incentivized review at all [Figure 6]. The Alexandrian Autograph Collection hotel has highest number of incentivized reviews (63) while Americas Best Value Inn places second position with 53 incentivized reviews. Hotels like Days Inn and Suites Albany, Galleria Park Hotel, French Market Inn, Hilton Garden Inn Orlando Airport, Hotel Diva etc. have no incentivized reviews.

The Variance graph [Figure 7] shows how number of reviews and incentivized reviews varied in our experiment.

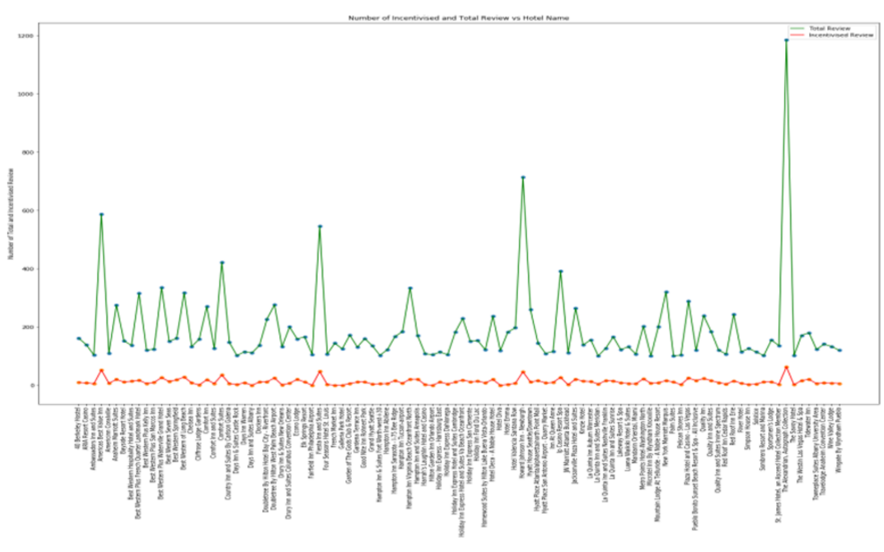

Figure 7. Incentivized Reviews with Respect to Total Reviews

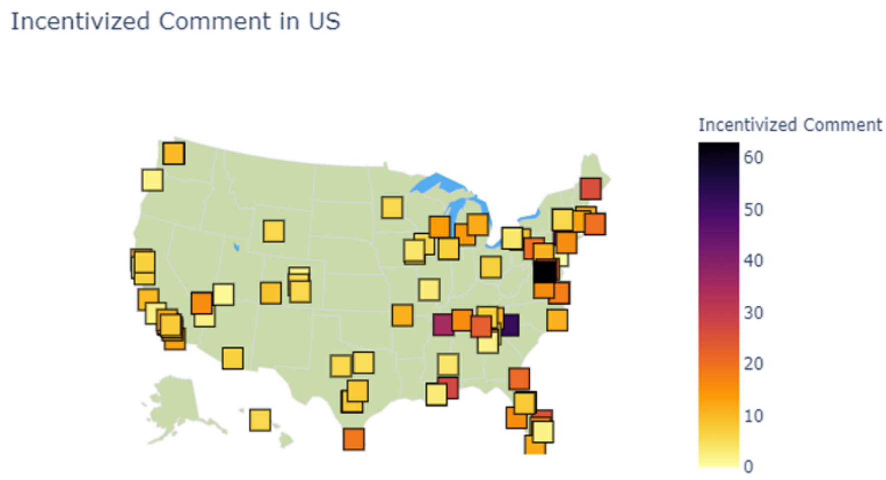

Figure 8. US Map with Incentivized Comments

Figure 8 shows incentivized comments in different locations of US. How incentivized comments vary between different cities in US are shown in the figure.

5. Conclusion

While booking hotel via online, reviews are crucial in deciding to purchase for most of the customers. Since customers have no physical interaction with the products, they fully rely on reviews thinking that they reflect real experiences of consumers. But incentivized or biased reviews often mislead customers who are going to book hotel. This paper analyses the online reviews on US hotels to identify incentivized reviews. Statistics of biased reviews in term of individual hotel and different cities of US are shown. The results have been validated using popular machine learning algorithms. People should be more aware about online frauds while booking hotels or intending to get other services or products. Supporting Information is available from the Wiley Online Library or from the author.

Acknowledgements

The researcher would like to thank the Department of Computer Science in Pabna University of Science and Technology for overall cooperation.

Conflict of Interest The authors declare no potential conflict of interest.

References

[1] Abdalraouf Hassan, Ausif Mahmood, presented at 2017 3rd International Conference on Control, Automation and Robotics,**2017** .

[2] Ana Costa, João Guerreiro, Sérgio Moro, Roberto Henriques,*Journal of Retailing and Consumer Services* **2019** ,47, 272-281.

[3] Bin Lin, Fiorella Zampetti, Gabriele Bavota, Massimiliano Di Penta, Michele Lanza, Rocco Oliveto, presented at *ACM/IEEE 40th International Conference on Software Engineering* ,**2018** ,.

[4] Daniel S. Kostyra, Jochen Reiner, Martin Natter, Daniel Klapper, *International Journal of Research in Marketing* **2016** , 33, 11-26 .

[5] Dina Mayzlin, Y Dover, J Chevalier, American Economic Review**2014** , 104, 2421-55.

[6] Dipak Gaikar, Bijith Marakarkandy, International Journal of Computer Science and Information Technologies**2015** , 6,2303-2313.

[7] Dirk Hovy and Anders Johannsens, presented at Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16), 2986–2989,**2016** .

[8] Doaa Mohey El-Din Mohamed Hussein, *Journal of King Saud University – Engineering Sciences* **2018** ,30, 330-338.

[9] Inyoung Chae, Andrew T. Stephen, Yakov Bart, Dai Yao, Marketing Science**2016** , vol. 36, 89-104.

[10] Michael Crawford, Taghi M. Khoshgoftaar, Joseph D. Prusa, Aaron N. Richter, Hamzah Al. Najada, *Journal of Big Data 2* **2015** , 23.

[11] M. Ghiassi, S. Lee, Expert Systems with Applications**2018,** 106, 197-216.

[12] Mohammad Masudur Rahman, Chanchal K Roy, and Raula G Kula, presented at *IEEE/ACM 14th International Conference on Mining Software Repositories (MSR),* 215-226,**2017** .

[13] Mohammed Salehan, Dan J. Kim, *Decision Support Systems***2016** *,* 81, 30-40.

[14] Natali Ruchansky, Sungyong Seo, and Yan Liu, presented in Proceedings of the 2017 ACM on Conference on Information and Knowledge Management. ACM, 797–806, **2017** .

[15] Oluwaseun Ajao; Deepayan Bhowmik; Shahrzad Zargari, presented at ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2507-2511, **2019** .

[16] Rajesh Piryani, Devaraj Madhavi, Vivek Kumar Singh, Information Processing & Management**2017** ,34,122-150.

[17] Sangkil Moon, Wagner A. Kamakura, *International Journal of Research in Marketing* **2017** *,* 34, 265-285.

[18] Soroush Vosoughi, Deb Roy, and Sinal Aral, Science**2018** , 359,1146–1151.

[19] Wei Wang, Hongwei Wang & Yuan Song, Journal of Experimental & Theoretical Artificial Intelligence**2016** , 29, 227-246.

[20] Wenqi Zhou, Wenjing Duan, Decision Support Systems**2015** ,76, 14-23.