Journal of Biological Engineering Research and Review, 2021; 8(2): 280-285
ISSN: 2349-3232 Online
Available online at www.biologicalengineering.in/Archive

Review Article

Web Data Mining: Sentiment Analysis of Amazon Product Reviews

Biswajit Biswas, Manas Kumar Sanyal, Tuhin Mukherjee

Department of Business Administration, University of Kalyani, West Bengal, India

***Corresponding author. E-mail:** biswajit.biswas0012@gmail.com **(B. Biswas)**

ARTICLE INFO:

Article History:
Received: 30/06/2021
Revised: 28/10/2021
Accepted: 11/12/2021
Published: 31/12/2021

Keywords:
Web-scraping; Data Mining; Feedback analysis; Review mining.

Citation: Biswas *et al.,* Web Data Mining: Sentiment Analysis of Amazon Product Reviews *J BiolEngg Res & Rev*, Vol. 8, Issue2, 2021

Abstract: Sentiment analysis on customer feedback in e-business is a rapid growing research area in the Business Intelligence System (BIS). A huge number of reviews numeric ratings and textual reviews (quantitative and qualitative) are shown on the e-market platforms on every product. These reviews are very helpful for the new buyers, sellers, retailers and manufacturers also. But it is very difficult or impossible to analysis this reviews data manually for business decision. Author proposed and developed a customer feedback or reviews analysis system that can extract the real time online review data from the e-marketing website and then analysis the sentiment of the extract reviews. Marketers and Customers can utilize review mining and feedback analysis, which have influenced the neighboring world by moving their belief on a particular product. Data manipulate in this research work are real-time product reviews (Galaxy S20 5G) collected from e-marketing site (amazon.in) through the web-scraping by python programming. The Author implemented a relative sentiment analysis of retrieved online reviews. This work provides feedback analysis of a mobile phone (smart phone) reviews dividing them into positive class, neutral class, and negative class. The result of this analysis has been visualized by a bar chart, word cloud, and numerical value, which can be used by a customer in decision-making before purchasing a new smart phone. The sellers can also use this system to expand their business.

INTRODUCTION

In 21st Century due to technological revolution, use of digital marketing/ e-commerce increased exponentially. Huge numbers of products are available in the e-market sites. It is very difficult for a customer to select a product from an embrace number of similar product. The online shopping style of customers is greatly affected by the development of online marketing sites. Reviews those exist on online marketing sites are used by a customer to make a right choice before purchasing decisions. The customer who has purchased a product can give their opinion on that product by giving rating, comments or both rating and comments.

A massive number of online reviews both ratings and comments are shown in e-marketing sites. It is very difficult for a customer to read these reviews and catch the truthful particulars about the product. To overcome these challenges and help the new buyers, feedback analysis is used. This technique is mainly used to extract the truth full information from the review's insight. This research work will help the new customers/buyers for making decision before selection of product from e-market sites.

RELATED WORK

In the rapid growing and spreading digital marketing most of the retailer and manufacturers want to sell their products through online marketing platforms. Customers can easily purchase their favorite products from these platforms and share their views/reviews on websites. To analyze the individual's subjective review, Fuzzy logic can be employed for a reliable meaningful insight [4]. The demand of e-commerce is rising and product reviews posted by customers give significant feedback for making potential customers and it influence the new buyers to take decision about their purchase. [7]

Sentiment Analysis is a procedure of determining the subjectivity and strength of polarity of specific comments. A survey confirms that 81% of e-market customers have done their on-line shopping by searching their favorite frequently. Manually feedback analysis is challenging to make a conclusion.

Today textual reviews available in the website are vigorously increased. There are numerous online marketing sites that permit customers to buy as well as post their comments for the procured products that results an incremental gathering of subjective reviews in natural

language (NL). To mine the inclusive sentiment or opinion polarity from altogether of them, sentiment analysis can be used for making a decision. In practical, manually these reviews analysis is impossible. Therefore, this efficient approach has an important role to solve these issues. [3]

A significant concept in fuzzy logic lies in the perception of semantic variables. Fuzzy set theory offers a traditional approach to model the intrinsic fuzziness between sentiment polarity classes. These characteristics make Fuzzy logic to identify sentiment classification of product reviews. [9]

Due to the continuous development of digital marketing platforms, traditional marketing and feedback system lost their importance buy that reason online reviews system take an important role in purchasing decisions. There is a problem in this system, how to analyze and find out the accurate information from these reviews. [1]

Feedback Analysis or Review mining is a process which analysis customer's feelings, attitudes, emotions and reactions towards certain entities. Data for this research work is together from Amazon.com. The Amazon reviews dataset analyzed accurately using python programming that classified the text reviews into positive, neutral or negative according to customer's textual feedback.[5]

The arena of sentiment analysis and review mining has an extensive possibility of research. It helps to invention the inclusive polarity of massive volume of dataset in no time and the result can be used for advance analysis for enhancement, and improvement of that specific product as well as the concern business [6]. Python based sentiment analysis results in the form of graphs and bar charts for easy visualization. [2]

Being motivated from the existing research the authors proposed a system that can be employed to predict, exact ratings and comments of a product. Sometimes the rating given by the customers does not furnish proper justification. By this research work, authors investigate sentiment analysis of a smart phone (Galaxy S20 5G) reviews from Amazon.in, that is valuable for the customers and the manufacturers also. Sentiment analysis analyzes the reviews and level it, i.e. 'better' and 'worse' sentiment as positive and negative respectively [8].

OBJECTIVE OF THE WORK

a. To identify the customers satisfaction level on a specific product.

b. To increase awareness about the product quality to the specific group of customers to enlarged sales.

c. To give related information from the existing users on a specific product for making purchases decision of a new buyer.

THE PROPOSED WORK

The authors proposed and developed a system to extract the real-time product reviews from the e-marketing website (amazon.in) of a particular product url. This review also analysis and visualized the inside sentiment with the help of python based NLTK, Beauty soup and word cloud. The confusion matrix also used to find out the accuracy of the sentiments and other statistical measures.

METHODS AND MODELS

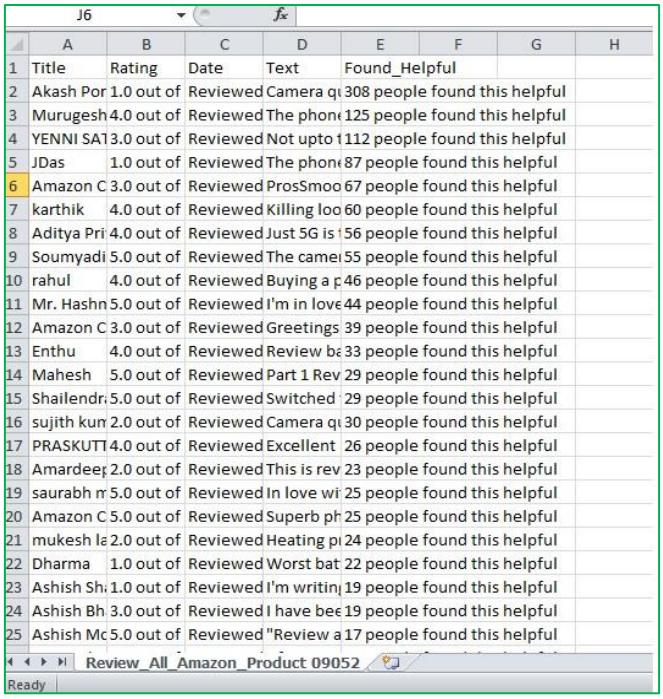
The number of reviews for a product is a dynamic set of information to be captured for sentiment analysis. The authors have devised the web-scraping program to exact the reviews for the selected amazon products. It's a python program using python package Beautiful Soup. It is a required python library to be imported for pulling data out of HTML & XML files. Parse Tree is created from the page source code to extract the product review data in hierarchy. Below are pre-requisites to run this review extraction program:

- ✓ Chrome Driver path based on the System OS & Directory path
- ✓ Selenium Web Driver to be imported
- ✓ Beautiful Soup to be imported

The inputs of this program are the review page links for the selected Amazon products. The outputs of the program are all the reviews shared by Amazon customers in a .csv file format. Amazon stores the heading title, rating, date, review text, found helpful flag for each review feedback shared by the customers for any product they purchased. The program is extracting all these fields in the .csv file. After extracting the customer feedback from Amazon, the authors have implemented sentiment analysis program using NLTK utilities available in Python. The input for this program is the extracted Amazon product reviews in .csv format. They have implemented the output visualization by bar chart and word cloud. Authors have used the confusion matrix to classify (polarize in to positive, negative and neutral words) the reviews and also measure the accuracy of the reviews.

RESULTS AND DISCUSSION

The developed system is working well, with the help of this programme we able to extract the e-marketing web site (amazon.in) product reviews. Here we insert Samsung **Galaxy S20 5G** smart phone link *(www.amazon.com/Samsung-Unlocked-Fingerprint-Recognition-Long-Lasting/dp/B082XXKZRC/ref=cm_cr_arp_d_product_top?ie=UTF8' driver.get(p_url))* in the input url section of the developed progamme on 5th May 2021, and run the programme. The output is automatically saved in name Review_All_Amazon_Product.csv' shown in Fig 1: From the Fig. 1: It is visible the reviewer's name, date of review given rating on the product, comment on this product and also it is shown that how many people found this is helpful for their decision making.



	A	B	C	D	E	F	G	H
1	Title	Rating	Date	Text	Found_Helpful			
2	Akash Por	1.0 out of	Reviewed	Camera q	308 people found this helpful			
3	Murugesh	4.0 out of	Reviewed	The phon	125 people found this helpful			
4	YENNI SA	3.0 out of	Reviewed	Not upto t	112 people found this helpful			
5	JDas	1.0 out of	Reviewed	The phon	87 people found this helpful			
6	Amazon C	3.0 out of	Reviewed	ProsSmoo	67 people found this helpful			
7	karthik	4.0 out of	Reviewed	Killing loo	60 people found this helpful			
8	Aditya Pri	4.0 out of	Reviewed	Just 5G is t	56 people found this helpful			
9	Soumyadi	5.0 out of	Reviewed	The camer	55 people found this helpful			
10	rahul	4.0 out of	Reviewed	Buying a p	46 people found this helpful			
11	Mr. Hashn	5.0 out of	Reviewed	I'm in love	44 people found this helpful			
12	Amazon C	3.0 out of	Reviewed	Greetings	39 people found this helpful			
13	Enthu	4.0 out of	Reviewed	Review ba	33 people found this helpful			
14	Mahesh	5.0 out of	Reviewed	Part 1 Rev	29 people found this helpful			
15	Shailendr	5.0 out of	Reviewed	Switched	29 people found this helpful			
16	sujith kun	2.0 out of	Reviewed	Camera q	30 people found this helpful			
17	PRASKUTT	4.0 out of	Reviewed	Excellent	26 people found this helpful			
18	Amardeep	2.0 out of	Reviewed	This is rev	23 people found this helpful			
19	saurabh m	5.0 out of	Reviewed	In love wi	25 people found this helpful			
20	Amazon C	5.0 out of	Reviewed	Superb ph	25 people found this helpful			
21	mukesh la	2.0 out of	Reviewed	Heating pi	24 people found this helpful			
22	Dharma	1.0 out of	Reviewed	Worst bat	22 people found this helpful			
23	Ashish Sh	1.0 out of	Reviewed	I'm writin	19 people found this helpful			
24	Ashish Bh	3.0 out of	Reviewed	I have bee	19 people found this helpful			
25	Ashish Mc	5.0 out of	Reviewed	"Review a	17 people found this helpful			

Review_All_Amazon_Product 09052

Ready

Fig 1: Website's extracted reviews in .csv format

In the second part of the developed programme it is need to insert the .csv file in the input section and it make the sentiment analysis for the extracted product reviews in the first section. In Fig 2: it is shown that the .csv file converted into a table.

Out[4]:

	Title	Rating	Date	Text	Found_Helpful
0	Akash Porwal	1.0 out of 5 stars	Reviewed in India on 22 March 2021	Camera quality is very poor, 108 MP camera i...	308 people found this helpful
1	Murugeshan Thevar	4.0 out of 5 stars	Reviewed in India on 22 March 2021	The phone is really great, I had purchased it ...	125 people found this helpful
2	YENNI SATYANARAYANA	3.0 out of 5 stars	Reviewed in India on 22 March 2021	Not upto the mark, compared to pro max only pr...	112 people found this helpful
3	JDas	1.0 out of 5 stars	Reviewed in India on 22 March 2021	The phone is awesome. Good camera and good dis...	87 people found this helpful
4	Amazon Customer	3.0 out of 5 stars	Reviewed in India on 22 March 2021	ProsSmooth displaySoundAmoled displaySome spec...	67 people found this helpful

Fig 2: Website's extracted reviews in a table

Out[22]:

	Text	Rating	Polarity_Rating	Review	Review_Processed
0	Camera quality is very poor, 108 MP camera i...	1	Negative	Camera quality poor 108 MP camera working 6 ...	Camera quality poor 108 MP camera work 6 MP ...
1	The phone is really great, I had purchased it ...	4	Positive	phone really great purchased first flash sale ...	phone really great purchase first flash sale l...
2	Not upto the mark, compared to pro max only pr...	3	Neutral	upto mark compared pro max pro better buy 2000...	upto mark compare pro max pro good buy 20000 v...
3	The phone is awesome. Good camera and good dis...	1	Negative	phone awesome Good camera good display use fea...	phone awesome Good camera good display use fea...
4	ProsSmooth displaySoundAmoled displaySome spec...	3	Neutral	ProsSmooth displaySoundAmoled displaySome spec...	ProsSmooth displaySoundAmoled displaySome spec...

Fig 3: Review process

Out[8]:

	Text	Rating
0	Camera quality is very poor, 108 MP camera i...	1
1	The phone is really great, I had purchased it ...	4
2	Not upto the mark, compared to pro max only pr...	3
3	The phone is awesome. Good camera and good dis...	1
4	ProsSmooth displaySoundAmoled displaySome spec...	3

Fig 4: Text review and correspondence rating

Out[9]: <AxesSubplot:xlabel='Rating', ylabel='count'>

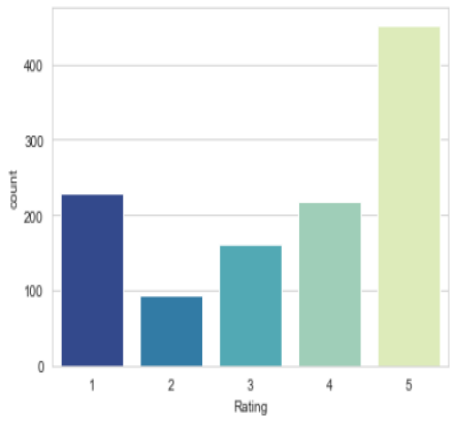


Fig 5: Representation of rating and count

In Fig 3: shows the review process, which is clearly shown that how the stop words and be verb, adverb, adjectives are removed from the text review sentences. It is also shown the text rating and polarity rating of each review.

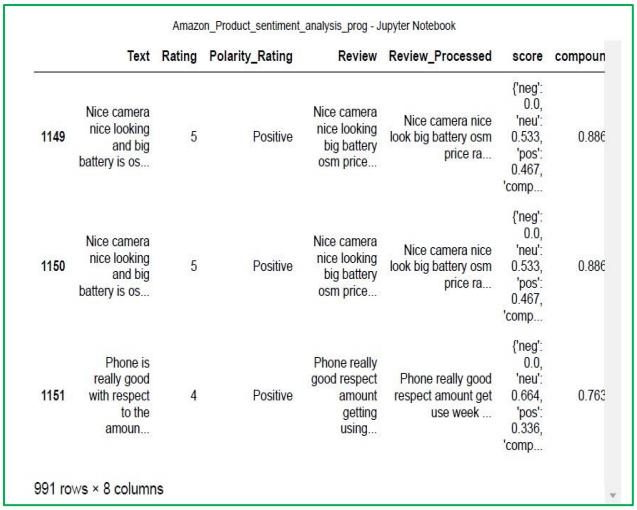
In Fig 4: the reviews are more clearly shown and the numerical reviews make correspondent with the text review.

In Fig 5: show a bar chart of the total review. This chart focused on the numerical rating i.e., X axis represent the rating (1 star, 2stars, 3stars, 4stars and 5stars) and the Y axis represent the count of the reviews.

Out[35]:

	Polarity_Rating	Review_Processed	score	compound
0	Negative	Camera quality poor 108 MP camera work 6 MP ...	{'neg': 0.383, 'neu': 0.617, 'pos': 0.0, 'comp...	-0.7351
1	Positive	phone really great purchase first flash sale l...	{'neg': 0.084, 'neu': 0.648, 'pos': 0.269, 'co...	0.9778
2	Neutral	upto mark compare pro max pro good buy 20000 v...	{'neg': 0.0, 'neu': 0.67, 'pos': 0.33, 'compou...	0.9153
3	Negative	phone awesome Good camera good display use fea...	{'neg': 0.158, 'neu': 0.71, 'pos': 0.133, 'com...	-0.3274
4	Neutral	ProsSmooth displaySoundAmoled displaySome spec...	{'neg': 0.0, 'neu': 0.734, 'pos': 0.266, 'comp...	0.7650

Fig 6: Polarity rating of text reviews

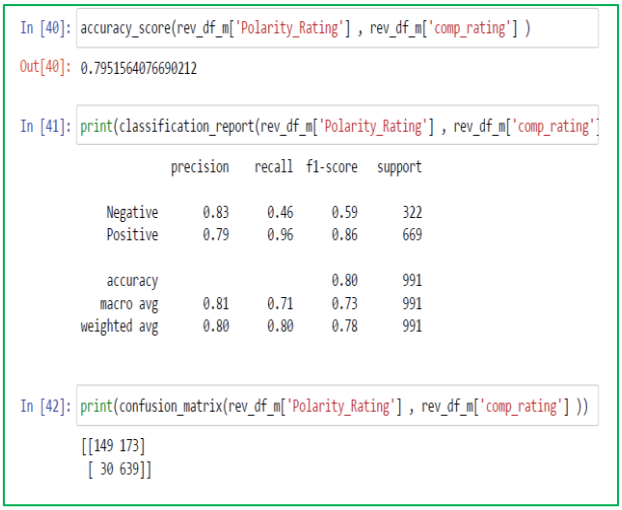


Amazon_Product_sentiment_analysis_prog - Jupyter Notebook

	Text	Rating	Polarity_Rating	Review	Review_Processed	score	compoun
1149	Nice camera nice looking and big battery is os...	5	Positive	Nice camera nice looking big battery osm price...	Nice camera nice look big battery osm price ra...	{'neg': 0.0, 'neu': 0.533, 'pos': 0.467, 'comp...	0.886
1150	Nice camera nice looking and big battery is os...	5	Positive	Nice camera nice looking big battery osm price...	Nice camera nice look big battery osm price ra...	{'neg': 0.0, 'neu': 0.533, 'pos': 0.467, 'comp...	0.886
1151	Phone is really good with respect to the amoun...	4	Positive	Phone really good respect amount getting using...	Phone really good respect amount get use week ...	{'neg': 0.0, 'neu': 0.664, 'pos': 0.336, 'comp...	0.763

991 rows × 8 columns

Fig 7: Compound polarity score



```
In [40]: accuracy_score(rev_df_m['Polarity_Rating'] , rev_df_m['comp_rating'] )
Out[40]: 0.7951564076690212

In [41]: print(classification_report(rev_df_m['Polarity_Rating'] , rev_df_m['comp_rating']
                precision    recall  f1-score   support

    Negative       0.83      0.46      0.59       322
    Positive       0.79      0.96      0.86       669

    accuracy                           0.80       991
   macro avg       0.81      0.71      0.73       991
weighted avg       0.80      0.80      0.78       991

In [42]: print(confusion_matrix(rev_df_m['Polarity_Rating'] , rev_df_m['comp_rating'] ))
[[149 173]
 [ 30 639]]
```

Fig 10: Accuracy and other statistical measure

In Fig 6: it is more clearly shown the total review process. First it shows the polarity rating (negative or positive) of the review then it analysis sentence wise and the part of the negative, positive and neutral sentiment of each review sentence. At last column it represents the total sentiment score of the review. From the Fig 7: it is observed that the compound sentiment and the polarity rating give the same information on each customer's review. If both the information (rating and text) not match then we may consider that the customer does not give accurate review on the said product.

Fig 8: Word cloud for positive reviews

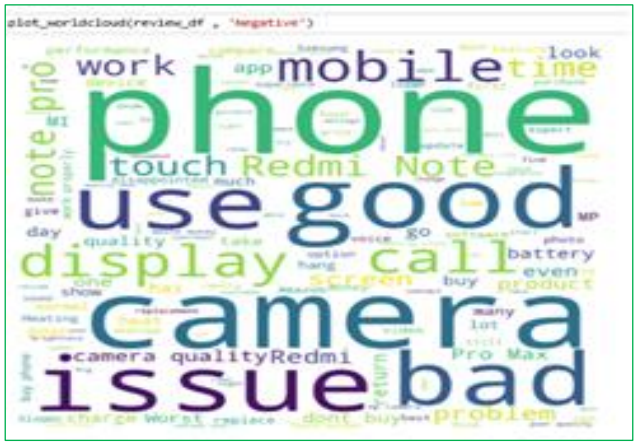


Fig 9: Word cloud for negative reviews

In the Fig 8: and Fig 9: shows the positive word cloud and negative word cloud. It is the visible result of the developed programme. In this word clouds the bold and big font words said the information gathered more strongly and small and normal font words comparatively gathered week information. These word clouds draw automatically, depending on the frequency of the words. Anyone can understand the exact on-site sentiment of the product review. In Fig 8: the positive word cloud said about the mention phone is good according to its battery, performance, display etc. and the Fig 9: the negative word cloud said that the mention phone is good but the camera of the mention phone is bad. If we analyze the FIGURE VIII and Fig 9: correspondingly then it is too clear that customers are compare the Samsung smart phone with Redmi smart phone. So, the manufacturers are needed to focus on camera and other features of Redmi. In Fig 10: authors show the accuracy of the sentiment analysis and other statistical measures. The confusion matrix shows the accuracy 0.795 for this particular product sentiment analysis. It also shows the precision, recall, f1-score and support in column on correspondence in rows, negative, positive, accuracy, macro avg. and weighted avg.

Review_All_Amazon_Product.csv' shown in Fig 1: From the Fig. 1: It is visible the reviewer's name, date of review given rating on the product, comment on this product and also it is shown that how many people found this is helpful for their decision making.

CONCLUSION

Customers' feedback analysis in e-business is a web-data mining which analyze their sentiments, attitudes, and emotions towards a particular product or service. The data of online product reviews for this research are collected from Amazon.com for a particular product over a predefined time interval. Uniqueness of this work is to find the overall polarity of enormous dataset in a few seconds with the help of a newly developed automated system.

In this research work Python based NLTK has been used for sentiment analysis of the comments. The newly developed application has been used to gain insights into user's

reaction to a specific product, based on the number of his/her ratings and comments. The application further provides focused semantic information such as positive feedback class, neutral feedback class and negative feedback class with reference customer's feedback. Results have been presented in form of graphs, bar charts and word cloud for easy visualization and understandable for a common people.

This analytical research work is helpful to depict the insight of, actual rating of a product, when customer's text review contradicts with numerical ratings. This insight of our experiment is beneficial for customers, Retailers, Competitors, and Manufacturers as well.

In a nutshell, this work is generic and can be used for any given product on a predefined data. In particular this experiment is done for the product model Samsung Galaxy S20 5G with reference to the review collected on 09.05.2021 using our newly developed python based NLTK augmented system software. The result suggests that customers recommend this product to future buyer with reference to value for money. However, the results suggest that manufacturer to improve the quality of charger for better market capitalization.

This empirical study is just a footstep in this research domain. Future researchers can explore the validity of our findings for a different product or with respect to a different e-business portal other than Amazon. The authors of this study expect the future researchers to explore the impact other types of feedback and the use of other techniques which have not been used in this paper due to paucity of time & other resource limitations.

ACKNOWLEDGEMENT

The authors wish to acknowledge and appreciate the University of Sheffield for making available a plethora of library resources that aided easy access to many of the closed sourced journals consulted in writing up this review.

CONFLICT OF INTEREST

Our article uses publicly available data across Ecommerce websites. This article does not contain any studies with human participants or animals performed by any of the authors. Further, all three authors are attached in the Department of Business Administration, University of Kalyani and consequently they used the infrastructure of their University to carry on this research activity. Consequently there is no conflict of interest involved in this case.

REFERENCES

1. Choudhary.M and Inventive Technology (ICSSIT), IEEE Xplore Part Number; CFP18P17-ART;ISBN:978-1-5386-5873-4
2. Dabas.C, Kaur.P, Gulati.N and Tilak.M,"Analysis of Comments on Youtube Videos using Hadoop", 5th International Conference on Image Information Processing (ICIIP), IEEE;ISBN:978-1-7281-0899-5/19/$31.00
3. Indhuja.K and Reghu Raj. P.C.2014,"Fuzzy Logic Based Sentiment Analysis of Product Review Documents",First International Conference on Computational Systems and Communications (ICCSC) 978-1-4799-6013-2/14/$31.00 _c 2014 IEEE, Trivandrum(2014),
4. Mhd. Altabbaa .T, Ayabakan. T and Ogrenci. A.S.2017," Fuzzy Logic for Decision Extraction from Product Reviews",International Conference on Energy, Communication, Data Analytics and Soft Computing (ICECDS-2017)
5. Pankag, Pandey.P, Muskan and Soni.N,"Sentiment Analysis on Customer Feedback Data: Amazon Product Reviews", International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, IEEE;ISBN:978-1-7281-0211-5/19/$31.00 2019
6. Rathee.N, Joshi.N and Kaur.J.2018,"Sentiment Analysis using Machine Learning Techniques on Pyhhon", Procedings of the Second International Conference on Intelligent Computing and Control Systems(ICICCS 2018), IEEE Xplore Complaint Part Number CFP18K74ART;ISBN 978-1-5386-2842-3,2018
7. Ravi.K and Ravi.V.2015, "A survey on opinion mining and sentiment analysis: Tasks, approaches and applications," Knowledge Based Systems, vol. 89, pp. 14-46, November 2015.
8. Sultana.N, Kumar.P,.Patra. M.R, Chandra S. and S.K. Alam. S.2019," Sentiment Analysis for Product review"; ICTACT JOURNAL ON SOFT COMPUTING, APRIL 2019, VOLUME: 09, ISSUE: 03, ISSN: 2229-6956 (ONLINE), DOI: 10.21917/ijsc.2019.0266.
9. Zadeh. L. A. *Calculus of Fuzzy Restrictions.*, Fuzzy Sets and their Applications to cognitive and decision processes, pages 1-40,1975.

About Authors

Mr. Biswajit Biswas, Faculty Member and Research Scholar in the Department of Business Administration, University of Kalyani. He has completed his B.Tech and MBA in Information Technology from the University of Kalyani, India. He has qualified the UGC-NET. He has five years research and teaching experience in Post Graduate and Under Graduate level in the domain of Business Administration. His expertise is about Business Intelligence, Web-data mining, Soft computing, Principle of Management, Strategic Management and Digital Marketing. He has published many research articles in International journal and conferences. He also visited many prestigious academic Institute like IIMA, IIMB, IIMC, IISC, AMITY University, KIIT, DUYTUN University Vietnam and many more in India and aboard. He has attached several teaching institutes including undergraduate & postgraduate levels as an examiner, paper setter and moderator.

Prof. Manas Kumar Sanyal, M. Tech. (Computer Science, University of Calcutta), Ph.D., Professor in the Department of Business Administration, University of Kalyani, India. He is the Life Member of Computer Society of India. Since December 2020, he is working as Vice-Chancellor of University of Kalyani. He was former Dean Faculty of Engineering, Technology & Management. His areas of research interest are Data Science, Big data, deep Learning and Business Intelligence. He has 27 years of teaching and research experiences. Till now seven Scholars have been awarded Ph.D. and five are pursuing. He has published more than hundred papers and three books from LAP Lambert, Germany.

Dr. Tuhin Mukherjee, Assistant Professor in The Department of Business Administration, University of Kalyani has a research experience of more than one decade and published many papers in national and international journals. He uses to teach in Post Graduate level in the domain of Business Administration. His expertise is about Business Analytics, Statistical Methods, Machine Intelligence, and Financial Management & Option Pricing. He already served several teaching institutes including undergraduate & postgraduate levels