JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Sentiment Analysis through Text Mining to rate the e-Commerce products based on the customer feedback

[1]Jyoti Verma, [2]Dr. Jaimin N Undavia
[1]Assistant Professor & Research student
[2]Associate Professor,
[1]ROFEL BBA & BCA College ,VApi
Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA) Faculty of Computer Science and Applications (FCA)
CHAROTAR UNIVERSITY OF SCIENCE & TECHNOLOGY (CHARUSAT)

[2] Smt. Chandaben Mohanbhai Patel Institute of Computer Applications (CMPICA)
Faculty of Computer Science and Applications (FCA)
CHAROTAR UNIVERSITY OF SCIENCE & TECHNOLOGY (CHARUSAT)

ABSTRACT

In recent time, passive or indirect feedback system evolve and become most popular to get rid of exact mind set of the peers towards the product or service offered by the provider. Sentiment analysis is the most widely used phenomenon for the same which mines the data, view, review, rating or sentence to predict the emotion of the said statement through different techniques like text mining or Natural Language Processing. Through such text mining we can classify the feedback into defined classes like "Happy", "Satisfied" or "Not Satisfied". Various tokens are getting segregated from the given feedback and further analysed to conclude the overall feedback. The advent use of e-commerce platforms now become a huge source of row data which is generated by the user or customer and these data is being used by the analysts to improve their service/product to achieve high level of customer satisfaction. To analysis of such huge data automatically, the field of sentiment analysis has turn up. The main aim of sentiment analysis is to identifying polarity of the data in the Web and classifying them. Sentiment analysis is text based analysis, but there are certain challenges to find the accurate polarity of the sentence.

KEYWORDS : Data mining , Sentiment analysis , opinion mining, machine learning, NLP, flair , pandas, Numpy

I. INTRODUCTION

Use of the internet has increased around the world like the fire in the forest and hence the popularity of the ecommerce website has increased. With the immense popularity of the online selling the users have become the customers of the online shopping. But with the online shopping the users are not able to physically see the product unless it is delivered so the product's reviews plays a vital role in the user's decision making to buy or not to buy the product. The user' vigorously thoroughly go through the reviews of the customers who have already purchased the product.

Data mining(Talib et al., 2016) finds the hidden information from the raw data and it does statistical analysis works Sentiment analysis also referred to as opinion mining is filtering the view or the emotions behind the test used in expression of the thoughts. The reviews of the customer have been a blessing in disguise for the online sellers as they are able to get the idea about the popularity or the failure of the product sold. Also the manufactures get the information about the enhancement of their product. The reviews have been mined to get such information and to do the mining text mining and NLP have always working together.

Reviews not only help built a positive or negative impression about a product or a service but also it can be found out whether the review is helpful or not(Vermeulen et al., 2017)

The customer reviews can give an insight into the products good features and its usefulness(Jack & Tsai, 2015). The reviews are not only restricted to only the product but also to the service. The (C. Cuizon et al., 2018) have used to text mining to get the information about the hotel based on the customers feedback given for the hotel.

The customer reviews can be used to get the competitor's business performance by analyzing the social media customer reviews, can convert the like thought of the customer to buy, enhance the business by enhancing the product, get the more likelihood of different product categories, addressing the problematic areas of the product or service.

The sentiment analysis(Rajput & Chauhan, 2019) of the tweets and the reviews can be done from which once can find when the user's reaction is positive or negative or neutral towards an event using the algorithms.

The rest of the paper follows as per the given order the section 2 is literature review, section 3 is the research problem, section 4 is methodology, section 5 is the findings and last section is the conclusions and future scope

II. LITERATURE REVIEW:

Author/Date	Focus/Topic/Question	Concept Theoretical Method	Context/Sample/Setting	Finding	Future Research
Jonathas G. D. Harb, R´egis Ebeling, Karin Becker [October 2019](Jonathas G. D. Harb, Regis Ebeling, 2019)	Sentiment analysis on the terrorist tweets	Data mining, machine learning supervised algorithms,LSTM (Long-Short Term Memory) and convolutional neural networks (CNN)	Social media tweets	The results of our analysis reveal that when terrorist events occur, a shift of emotion towards anger, sadness, and fear can be noticed .Analysis proved that Age and gender also affect and that age is the most influential factor .	
Wu Hea,Shenghua Zha , Ling Li (He et al., 2013)	Social media competitive analysis and text mining: A case study in the pizza industry	Text mining , SPSS Clementine's linguistic methods (extracting, grouping, indexing, etc.)	Facebook, Twitter	The users have used the social media to give their views on pizza and this views have proved to be useful for the sellers to understand their customer requirements and also develop good relation with their customers	finding innovative ways to turn businesses' social media fans from "like" to "buy"
L. Jack and Y.D. Tsai (Jack & Tsai, 2015)	Using Text Mining of Amazon Reviews to Explore User-Defined Product Highlights and Issues	Text mining	Amazon product review	The users use specific words for their positive or negative or neutral reviews	Deep machine learning process can be applied to do predictive analysis
Subhasis Dasgupta Kalyan Sengupta (Dasgupta & Sengupta, 2016)	Analyzing Consumer Reviews with Text Mining Approach: A Case Study on Samsung Galaxy S3	Text mining, word cloud, document clustering, association rule mining		The reviews are of the specific product and they suggest positive aspect of the mobile	This information can be further used for the up gradation of the product
Athira K, Dhanya M, Sinita S Ashok (Athira K, May 2020)	Application of Text Mining on Customer Buying Pattern of Cosmetic Products [March 2020] [tm14]	Text mining, classification	Ecommerce websites	The authors have identified that specific set of words which are used frequently by the customers	N/A

				to describe the products likeability. And the customers are inclined towards the purchase of the organic product for the cosmetics	
IEEE 20th Conference on Business Informatics (Chanwisitkul et al., 2018)	The Reason behind the Rating: Text Mining of Online Hotel Reviews	Text mining	Customer reviews form the hotel's website,social media posts	Assists the management to take top level decisions for providing better facilities	
Dashrath Mane, Dr. Prateek Srivastava [2020] (Mane & Srivastava, 2020)	Netnography and Text Mining to Understand Perceptions of Indian Travellers using Online Travel Services				
V. Uma Devi ,Dr. Vallinayagi V[Feb 2019](*PRODUCT REVIEW SENTIMENT ANALYSIS-A SURVEY*, 2019)	PRODUCT REVIEW SENTIMENT ANALYSIS – A SURVEY	Sentiment Analysis,Multilayer precepton, Support Vector Machine and Navie Bayes	Customer reviews of beauty products from amazon.com	Studied the results of the three algorithms Naïve bayes, SVM and MLP. And proved that MLP gave better result than Naïve Bayes & SVM	More algorithms can be tested.
Snigdha Dixit,Santosh Kr [May 2016](Dixit & Kr, 2016)	Collaborative Analysis of Customer Feedbacks using Rapid Miner	Rapid miner	Customer review on 3 different types of mobile phone Samsung, nokia lumia and sony experia	The authors found that Samsung is mostly purchased as compared to other two mobiles	The study can be extended to other mobile brands also and the study criteria can be analysed.
Arjun Mukherjee,Bing Liu,Natalie Glance(Mukherjee et al., 2012)	Spotting Fake Reviewer Groups in Consumer Reviews	Group Spam Behavior Indicators: GSRank	Customer reviews from amazon.com	The authors have developed the algorithm to detect the group spammers who give the fake reviews	

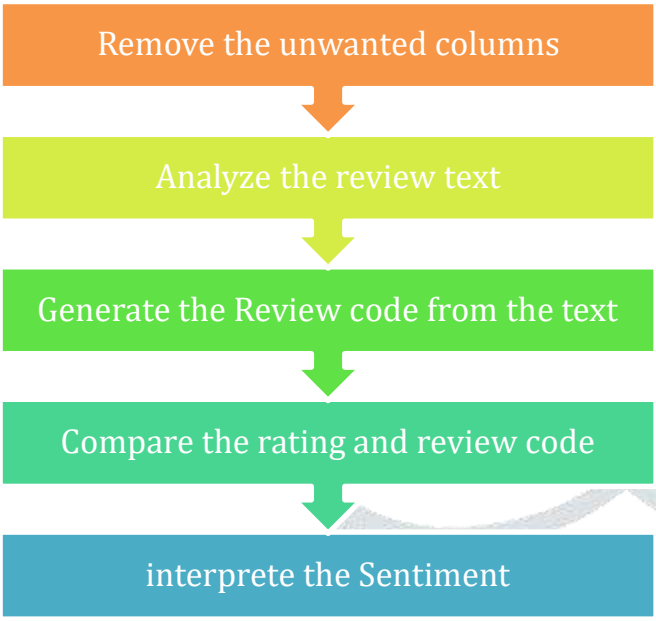
III. Research problem

The reviews of the customer are in the form of the unformatted text and sometimes the text is written in the natural language also so it must be converted to a form on which the research can be done and desired output can be obtained.

Herein the reviews of the electronic products is obtained from the amazon website and the collected review data is analyzed to understand the positive or the negative sentiments embedded in the words used by the customer.

To actually work with the data the textual data needs to be cleaned and then analyzed and to work in this direction the following steps will be considered.

Fig: Steps for finding the codes



IV. METHODOLOGY

Dependent variable Based on the dataset obtained from the amazon.com site the data values can be classified as dependent and independent variables

Independent variables:

1. Brand
2. Manufacturer
3. Review title
4. Review text
5. Rating
1. Sentiment status
2. Review code

Sentiment status can be:

The value of this variable will depend on the review text and rating

The sentiment will be positive if the value of title and rating are found to match

The sentiment will be negative if the value of the review text and rating differ.

Review code is a :

The code is generated by parsing the review text and comparing the individual words with predefined list of words representing the strong or weak review

Review text can be analyzed using the following criteria

- A review code can be generated by interpreting the review phrases and rating
- If review has phrases like 'excellent',' love',' loved', 'loves', 'best', 'fantastic', 'great' , 'recommend', 'useful', 'works', 'awesome' review code can be 5
- If review has phrases like 'nice', 'satisfied', 'good', 'better' than review code can be 4
- If review has phrases like 'inexpensive', 'like', 'liked', 'ok', 'fine' than review code can be 3
- if review has phrases 'expensive', 'costly', 'disappointed', 'bad', 'worse', 'worst','waste','useless','satisfactory','poor','terrible','worthless' review code can be 1

Sentiment analysis

If the review code is between 2 to 5 and the rating is also between 2 to 5 than positive (1) sentiment otherwise negative (0) sentiment

Comparison can be done once the sentiment code is generated and classification of the product as "good or bad " and "recommend or not"

Comparison of the values of the sentiment based on the reviews and rating. The values which are given is been produced by the model or not

Python has a rich set of libraries which are like the boon to a programmer for performing the task. The libraries which are used in this work are the pandas, flair for fetching the approximate value of the sentiment from the reviews collected.

The python libraries like:

Pandas

Flair: It is a simple framework for state-of-the-art NLP. It is a very powerful library which is developed by Zalando Research. The Flair framework is built on top of PyTorch[1].

It provided various functionalities such as:

pre-trained sentiment analysis models,

text embeddings, NER etc

Fig: Code for finding the value of sentiment

```
for i in range(0,df.shape[0]):
      text=df['reviews.text'][i]
      sentence = flair.data.Sentence(text)
      sentence.to_tokenized_string()
      model.predict(sentence)
      if sentence.get_labels()[0].value=='POSITIVE':
        v=1
      else:
        v=0
      df.at[i,'senti.status']=v
      df.at[i,'senti.value']=sentence.get_labels()[0].score
```

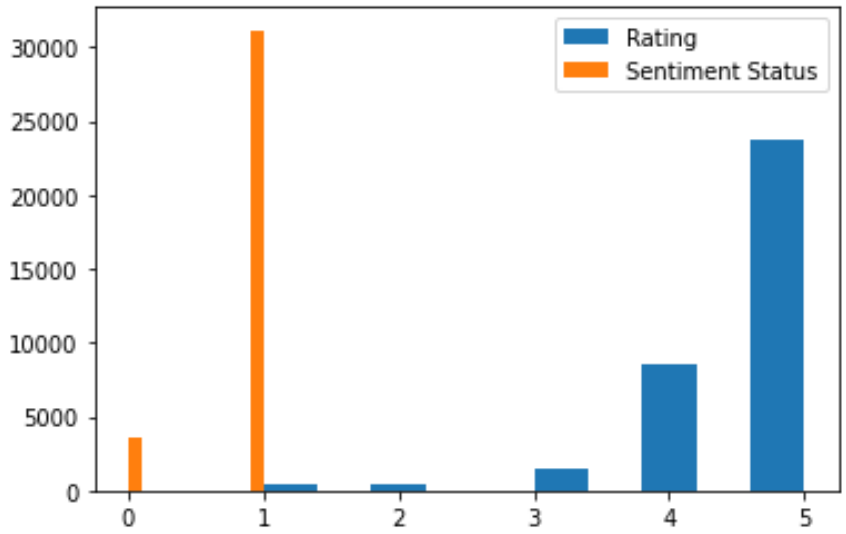
V. FINDINGS

The reviews written by the customers are parsed using the Flair library and it gives the same numerical values as rated by the user.

Fig: Result of running the algorithm

	categories	reviews.rating	senti.status	senti.value
0	Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Computers & Tablets	5	1	0.990324
1	Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Computers & Tablets	5	1	0.989165
2	Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Computers & Tablets	5	1	0.999079
3	Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Computers & Tablets	4	1	0.997476
4	Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Computers & Tablets	5	1	0.998628
5	Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Computers & Tablets	5	1	0.999918
6	Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Computers & Tablets	4	1	0.99647
7	Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Computers & Tablets	5	1	0.993557
8	Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Computers & Tablets	5	1	0.998892
9	Electronics,iPad & Tablets,All Tablets,Fire Tablets,Tablets,Computers & Tablets	5	1	0.957783
	Electronics,iPad & Tablets,All Tablets,Fire			

The value of the review rating and the value obtained from the sentiment analysis can be plotted in the graph as:



VI. CONCLUSION AND FUTURE SCOPE

The language used by the customers has certain common phrases and words which are used by them to express their feelings towards the product and they use it often. And these reviews can give an insight into the usefulness of the product. These reviews are available freely in in very large quantity but the information is unstructured and if converted to numerical values than it can easily be analyzed using the libraries available in machine learning and deep learning.

This paper has considered the review of various papers which have implemented the sentiment analysis on the customer feedback for either getting the future customers and improvisation of the product. The paper has used the flair library of the python to do the conversion of the text to numbers for analyzing the customer review.

VII. REFERENCES

[1] https://www.geeksforgeeks.org/flair-a-framework-for-nlp/

[2] https://neptune.ai/blog/sentiment-analysis-python-textblob-vs-vader-vs-flair

[3] C. Cuizon, J., Lopez, J., & Rose Jones, D. (2018). Text Mining Customer Reviews for Aspect-Based Restaurant Rating. *International Journal of Computer Science and Information Technology*, *10*(06), 43–52. https://doi.org/10.5121/ijcsit.2018.10605

[4] Chanwisitkul, P., Shahgholian, A., & Mehandjiev, N. (2018). The reason behind the rating: Text mining of online hotel reviews. *Proceeding - 2018 20th IEEE International Conference on Business Informatics, CBI 2018*, *1*, 149–157. https://doi.org/10.1109/CBI.2018.00025

[5] Dasgupta, S., & Sengupta, K. (2016). Analyzing Consumer Reviews with Text Mining Approach. *Paradigm*, *20*(1), 56–68. https://doi.org/10.1177/0971890716637700

[6] Dixit, S., & Kr, S. (2016). Collaborative Analysis of Customer Feedbacks using Rapid Miner. In *International Journal of Computer Applications* (Vol. 142, Issue 2).

[7] He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, *33*(3), 464–472. https://doi.org/10.1016/j.ijinfomgt.2013.01.001

[8] Jack, L., & Tsai, Y. D. (2015). Using Text Mining of Amazon Reviews to Explore User-Defined Product Highlights and Issues. *International Conference of Data Mining DMIN'15*, *July*, 92–98.

[9] Jonathas G. D. Harb, Regis Ebeling, K. B. (2019). Exploring deep learning for the analysis of emotional reactions to terrorist events on Twitter. *Journal of Information and Data Management*, *10*(2), 97–115. https://periodicos.ufmg.br/index.php/jidm/article/view/14711/16364

[10] Mane, D., & Srivastava, P. (2020). Netnography and Text Mining to Understand Perceptions of Indian Travellers using Online Travel Services. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 11, Issue 9). www.ijacsa.thesai.org

[11] Mukherjee, A., Liu, B., & Glance, N. (2012). Spotting fake reviewer groups in consumer reviews. *WWW'12 - Proceedings of the 21st Annual Conference on World Wide Web*, 191–200. https://doi.org/10.1145/2187836.2187863

[12] *PRODUCT REVIEW SENTIMENT ANALYSIS-A SURVEY*. (2019). www.jetir.org

[13] Rajput, N., & Chauhan, M. S. (2019). International Journal of Computer Science and Mobile Computing Analysis of Various Sentiment Analysis Techniques. *International Journal of Computer Science and Mobile Computing*, *8*(2), 75–79. www.ijcsmc.com

[14] Talib, R., Hanif, M. K., Ayesha, S., & Fatima, F. (2016). Text Mining: Techniques, Applications and Issues. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 7, Issue 11). www.ijacsa.thesai.org

[15] Tseng, Y. H., Lin, C. J., & Lin, Y. I. (2007). Text mining techniques for patent analysis. *Information Processing and Management*, *43*(5), 1216–1247. https://doi.org/10.1016/j.ipm.2006.11.011

[16] Vermeulen, I. E., Seegers, D., Singh, J. P., Irani, S., Rana, N. P., Dwivedi, Y. K., Saumya, S., Kumar Roy, P., Liu, S. Q., Mattila, A. S., Editors, D. R. F., Duan, W., Gu, B., & Whinston, A. B. (2017). Predicting the "helpfulness" of online consumer reviews. *Journal of Business Research*, *70*(4), 123–127. http://link.springer.com/10.1007/978-3-319-44263-1%0Ahttp://dx.doi.org/10.1016/j.tourman.2008.04.008