International Journal of Computational Intelligence Research
ISSN 0973-1873 Volume 13, Number 1 (2017), pp. 11-22

http://www.ripublication.com

Impact of Social Media and Customer Feedback on Online Auctions

Sandeep Kumar[1] and Rahul Rishi[2]

[1,2]CSED (UIET), MD University, Rohtak, India.

Abstract

Social associations are important to be known for making sense of buying selling process, the association amongst social and business systems has been widely investigated on a global online scale. In current scenario the localization impact of social events are greatly reflected through the available social media platforms. These platforms not only helps a customer to express free opinions but they also guide or motivate thousands of other buyers or bidders in making corresponding decisions. In this experiment, we focus about consideration on how an individual or group's business exchanges are motivated from the prevailing social events. Utilizing product specific scores and available open reviews, we measured the value of trust. By combing the two we have presented a framework on integrated approach. The implementations work shows the experiment setup for the same. How much will a buyer pay for product or service dependent on trust on upon seller or the trust built-up locally. To consider the more focused and real time opinion a multidimensional evolutionary analysis model have been shown. This paper also reveal the factors underlying the cost determination for an online auctions or buying patterns. At the end of paper, the graphs displaying social media trends have also been shown. These visualizations definitely improves one's understanding about a product or service.

Keywords: Online Auctions; Trust Mechanism; End Price; Data Mining; Software Agents; Clustering; Recommender System, Price Determinants.

1. INTRODUCTION

Online auctions began their existence in the mid 1990's when eBay the first website of this type, was began in California in 1995. Amongst number of famous online auction are Amazon.Com, Godaddy, Yahoo!, uBid and plenty of others. At the moment eBay is the largest and one of the fastest developing on-line public sale sites with consolidated net revenues of virtually manifold 2015 [12].

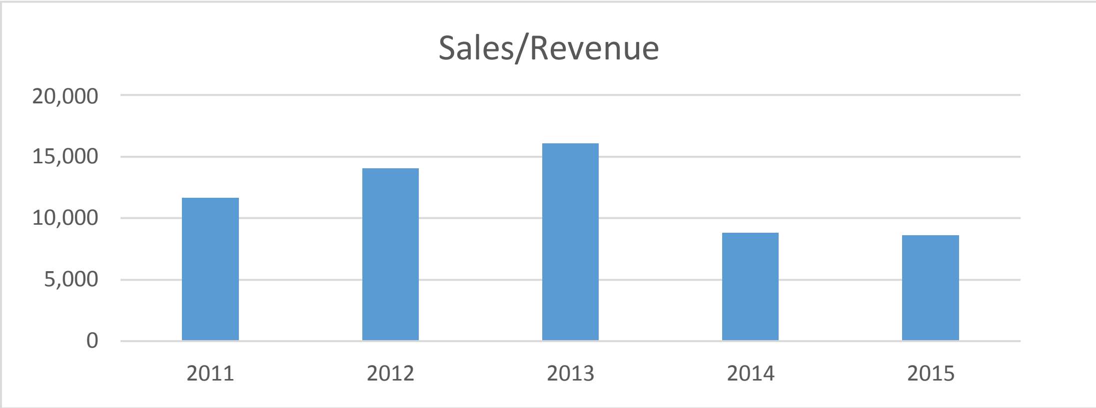


Figure 1. Following is the revenue generated by e-bay in different years.

As the digital markets proceed to grow at very rapidly, their status and availability of giant amounts of knowledge researchers' awareness from exclusive fields reminiscent of economics, marketing and some knowledge-associated spheres.

Social Media as a representative of local parameters and its impact on price determination:

This media achieves an essential position in uncovering the emotions of people about the exact problem or event. In Twitter and Facebook number of humans who posts on Facebook or who Tweet on Twitter may be very significant. Now the essential objective here is how the information can also be extracted from such social media platforms. This becomes much more difficult as the data from specific platforms tends to have unexpected characteristics. For illustration, it may be the number of shares or likes within the case of Facebook even as on Twitter it may be number of Tweets and so on. Even if go to the subsequent degree like in case of Twitter there are specified key terms that end up enormous. These small-small factors in datasets aid in making the conclusion as highly accurate [9].

For example the following is the list showing such factors that can

i. Specific keywords like good, bad, great etc. used by people.
ii. No. of Tweets in Twitter
iii. Who responds to whom analysis

iv. Pin Interest
v. Frequency of words used
vi. Any other links shared etc.
vii. Image counts
viii. No. of comments in Facebook
ix. No. of likes in Facebook
x. Shares in Facebook
xi. Google +1

There are many more depending on the analysis requirements. The certain unique fragments of data that are being collected by these social media platforms e.g. Facebook collects 63 different pieces of data elements [9].

2. PROPOSED MODEL

The proposed model proceeds through two phases:

1. First we follow the social media trend of an item.
2. Then we collect the review rating for that particular item auction.
3. By having a score from above two. We consider the relevance and social media impact on final price of auction.

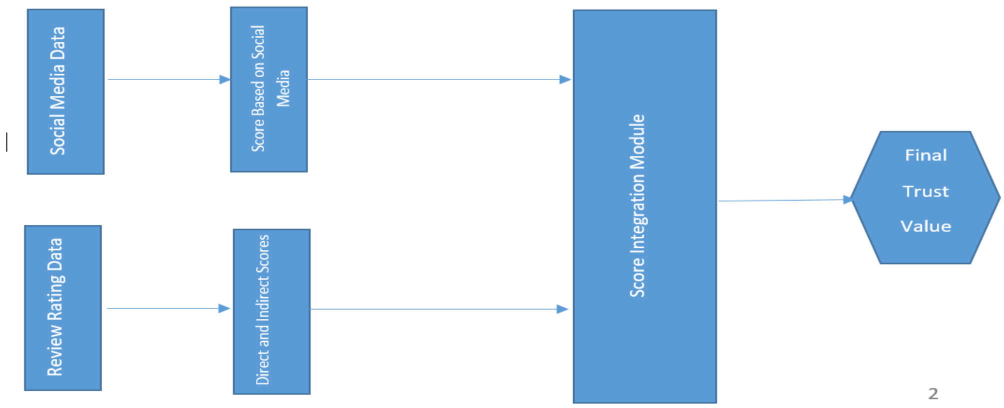


Figure 2: The figure shows the phases of trust building.

In Part-A Social media platform “Twitter” has been chosen for reading the local trend analysis. For this the location and in Part-B customer reviews has been used for getting information about how a customer thinks about a product. The important point to note here is that in Part-B multimode analysis has been performed i.e. other than the direct score provided by a user, the factors like timing of the review, the certified user and the comment text that user have provided etc. have been considered for predicting about the product.

After this Part-A and Part-B has been combined to give an integrated score. So the proposed model here not only shows the social media analytics but it also analyses the customer opinion from different aspects. In totality it gives a comprehensive image building frame of a particular product that is being sold by a seller online.

2.1 Framework for Social Media i.e. Twitter Data Collection:

Here Twitter has been considered for the main data source. The reason behind this is that it has been one of the most largely used social media in recent years and the number is growing at a very fast speed. Some of the amazing facts that strengthen its usage as representative of people's opinion are [6]:

1. Number of People having account on Twitter are about 1.3 billion.
2. Active daily users in Twitter are more than 100 million.
3. On Twitter more than 6000 tweets are available. Sometimes it is more than 100K.
4. More than 80% users are using mobile phone to have access of the same.

And one more important fact is that as per a survey by Twitter more than 54% user had been influenced in their decision after seeing the brand mentioned in the Tweets [7]. So from the above facts it would be wrong that in today's era Twitter can be a good representative of people's opinion.

In this framework the following modules have been used:

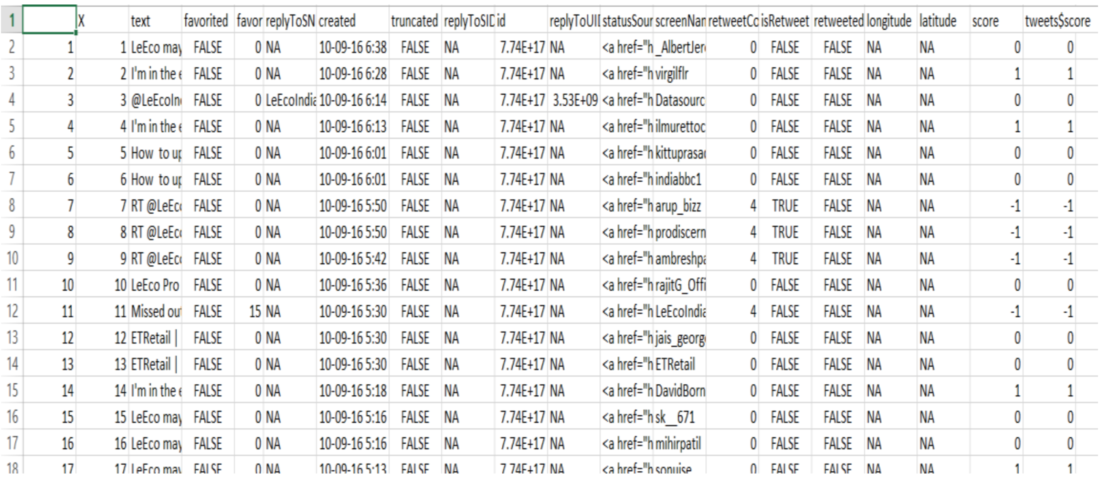
1. Data Collection:

Algo: Data_collection

Step1.:Linking the Twitter API.

Step2.:Keyword based search.

Step3.:Storing Tweets.



1		X	text	favorited	favor	replyToSN	created	truncated	replyToSID	id	replyToUI	statusSour	screenNan	retweetCo	isRetweet	retweeted	longitude	latitude	score	tweets$score
2	1	1	LeEco may	FALSE	0	NA	10-09-16 6:38	FALSE	NA	7.74E+17	NA	<a href="h	_AlbertJer	0	FALSE	FALSE	NA	NA	0	0
3	2	2	I'm in the	FALSE	0	NA	10-09-16 6:28	FALSE	NA	7.74E+17	NA	<a href="h	virgilflr	0	FALSE	FALSE	NA	NA	1	1
4	3	3	@LeEcoIn	FALSE	0	LeEcoIndia	10-09-16 6:14	FALSE	NA	7.74E+17	3.53E+09	<a href="h	Datasourc	0	FALSE	FALSE	NA	NA	0	0
5	4	4	I'm in the	FALSE	0	NA	10-09-16 6:13	FALSE	NA	7.74E+17	NA	<a href="h	ilmurettoc	0	FALSE	FALSE	NA	NA	1	1
6	5	5	How to up	FALSE	0	NA	10-09-16 6:01	FALSE	NA	7.74E+17	NA	<a href="h	kittuprasa	0	FALSE	FALSE	NA	NA	0	0
7	6	6	How to up	FALSE	0	NA	10-09-16 6:01	FALSE	NA	7.74E+17	NA	<a href="h	indiabbc1	0	FALSE	FALSE	NA	NA	0	0
8	7	7	RT @LeEc	FALSE	0	NA	10-09-16 5:50	FALSE	NA	7.74E+17	NA	<a href="h	arup_bizz	4	TRUE	FALSE	NA	NA	-1	-1
9	8	8	RT @LeEc	FALSE	0	NA	10-09-16 5:50	FALSE	NA	7.74E+17	NA	<a href="h	prodiscern	4	TRUE	FALSE	NA	NA	-1	-1
10	9	9	RT @LeEc	FALSE	0	NA	10-09-16 5:42	FALSE	NA	7.74E+17	NA	<a href="h	ambreshpa	4	TRUE	FALSE	NA	NA	-1	-1
11	10	10	LeEco Pro	FALSE	0	NA	10-09-16 5:36	FALSE	NA	7.74E+17	NA	<a href="h	rajitG_Offi	0	FALSE	FALSE	NA	NA	0	0
12	11	11	Missed ou	FALSE	15	NA	10-09-16 5:30	FALSE	NA	7.74E+17	NA	<a href="h	LeEcoIndia	4	FALSE	FALSE	NA	NA	-1	-1
13	12	12	ETRetail |	FALSE	0	NA	10-09-16 5:30	FALSE	NA	7.74E+17	NA	<a href="h	jais_georg	0	FALSE	FALSE	NA	NA	0	0
14	13	13	ETRetail |	FALSE	0	NA	10-09-16 5:30	FALSE	NA	7.74E+17	NA	<a href="h	ETRetail	0	FALSE	FALSE	NA	NA	0	0
15	14	14	I'm in the	FALSE	0	NA	10-09-16 5:18	FALSE	NA	7.74E+17	NA	<a href="h	DavidBorn	0	FALSE	FALSE	NA	NA	1	1
16	15	15	LeEco may	FALSE	0	NA	10-09-16 5:16	FALSE	NA	7.74E+17	NA	<a href="h	sk__671	0	FALSE	FALSE	NA	NA	0	0
17	16	16	LeEco may	FALSE	0	NA	10-09-16 5:16	FALSE	NA	7.74E+17	NA	<a href="h	mihirpatil	0	FALSE	FALSE	NA	NA	0	0
18	17	17	LeEco may	FALSE	0	NA	10-09-16 5:13	FALSE	NA	7.74E+17	NA	<a href="h	sonuise	0	FALSE	FALSE	NA	NA	1	1

Figure 3: Screenshot of Data Collected.

2. Data Cleansing:

Algo: Data_Clean

Step1.:Convert to datafreame

Step2.:Remove

Step3.:Remove stopwords

Step4.:Remove words with frequency.

3. Scoring The Sentiments:

Step1.:Load positive and negative files

Step2. Input to score_sentiment algorithm

Score_sentiment(positive,negative, tweets)

Traverse(1 to n)

For each negative tweet

Assign sentiments to words

Score= Positive Sentiments-negative tweets

Store(Score.tweets)

Result (tweet.df)

4. Predictive Model:

Splitting the Data and Finalizing the values for further analysis are performed here [10].

Algo. Predictive_Model

Step1.:Load (df)

Step2.:Split between trininig & test data.

Step3.:Train set-Baseline, Cart, RF

Step4.:Test set-Baseline, Cart, RF

Figure below shows the predictive analysis of the experimental setup using printcp(tweetCart):

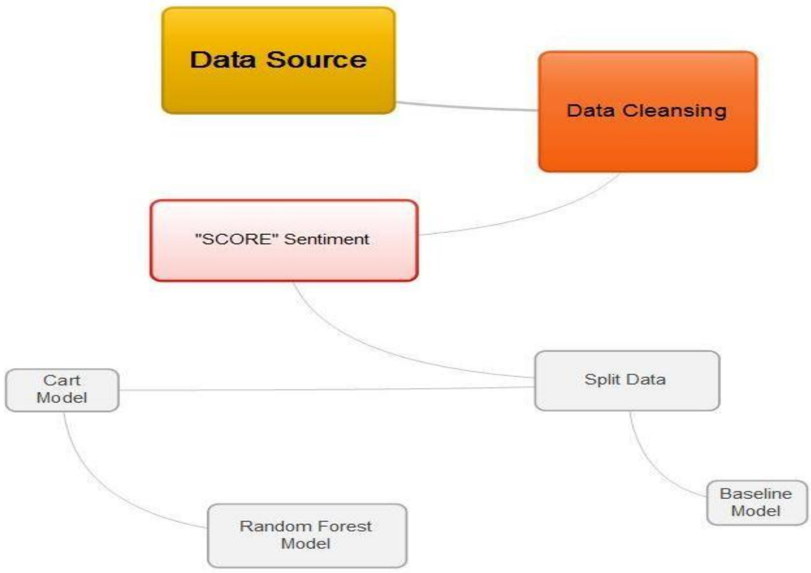


Figure 5. Predictive Analysis based on given input using the sentiment scores.

Phase 2. Analysing the reviewal based Trust Mechanism:
Next we have analysed how trust building can be performed in our model:
Utilizing seller scores and review ability, we then measure a cost of trust.
It is very significant to note that many a time this is not possible for a user to express his/her opinion freely. So we need some additional model dependent factors. These factors may come by combining multiple aspects like product description, shipment cost communication and handling or packaging issues etc. Here we will generate a trust score for a particular product.

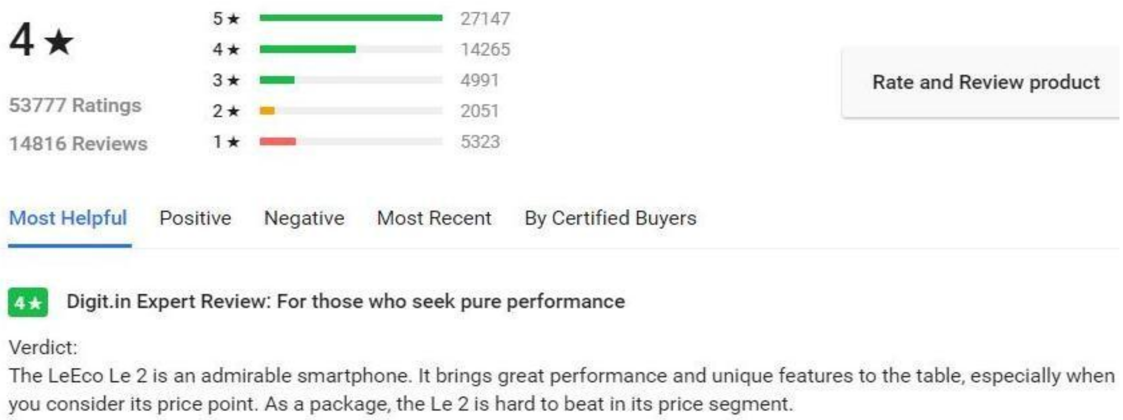


Figure 6. Shows Customer Review on a website.

In figure above the flipkart rating system has been shown. There are number of factors that are of significance, these have been discussed in the section below.
In figure below the proposed model for trust evaluation has been shown. It has following modules:

1. Dataset Building Module

For building the review analysis dataset, the reviews on "LeEco" have been chosen. The reason for the same is that a large number of reviews i.e. 6761 and a total star rating of 18899 is available on the internet (Source: www.Flipkart.com). The collection of reviews has been performed using a Java based comment scrapper. This has been highlighted in the experiment section of Part-B. After using the same we have obtained a fetched review data, which can be further used for analytics purpose. For dataset building the code for comment scrapper in Python has been written and the sample for the same is shown below:

The data collected from the above code was obtained in form of following format.

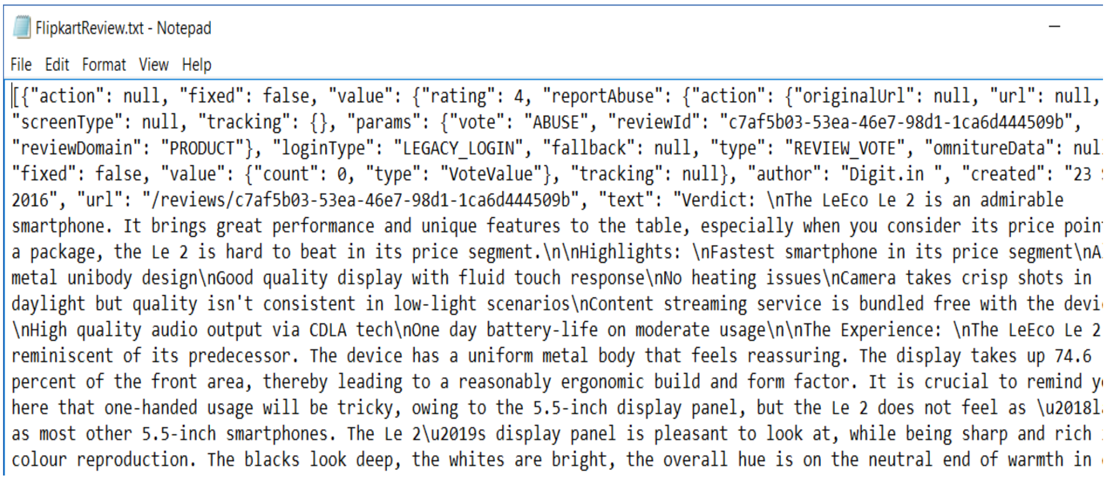

FlipkartReview.txt - Notepad
File Edit Format View Help
[{"action": null, "fixed": false, "value": {"rating": 4, "reportAbuse": {"action": {"originalUrl": null, "url": null,
"screenType": null, "tracking": {}, "params": {"vote": "ABUSE", "reviewId": "c7af5b03-53ea-46e7-98d1-1ca6d444509b",
"reviewDomain": "PRODUCT"}, "loginType": "LEGACY_LOGIN", "fallback": null, "type": "REVIEW_VOTE", "omnitureData": nul
"fixed": false, "value": {"count": 0, "type": "VoteValue"}, "tracking": null}, "author": "Digit.in ", "created": "23
2016", "url": "/reviews/c7af5b03-53ea-46e7-98d1-1ca6d444509b", "text": "Verdict: \nThe LeEco Le 2 is an admirable
smartphone. It brings great performance and unique features to the table, especially when you consider its price poin
a package, the Le 2 is hard to beat in its price segment.\n\nHighlights: \nFastest smartphone in its price segment\nA
metal unibody design\nGood quality display with fluid touch response\nNo heating issues\nCamera takes crisp shots in
daylight but quality isn't consistent in low-light scenarios\nContent streaming service is bundled free with the devi
\nHigh quality audio output via CDLA tech\nOne day battery-life on moderate usage\n\nThe Experience: \nThe LeEco Le 2
reminiscent of its predecessor. The device has a uniform metal body that feels reassuring. The display takes up 74.6
percent of the front area, thereby leading to a reasonably ergonomic build and form factor. It is crucial to remind y
here that one-handed usage will be tricky, owing to the 5.5-inch display panel, but the Le 2 does not feel as \u2018l
as most other 5.5-inch smartphones. The Le 2\u2019s display panel is pleasant to look at, while being sharp and rich
colour reproduction. The blacks look deep, the whites are bright, the overall hue is on the neutral end of warmth in

Figure 8. Data Collected from a website.

Now this data is in Jason flat format. Further we compiled this data and converted in to .csv format. The columns that have been used for analysis are:

1. Rating
2. Name of the author (Converted to encoded form)
3. Date of creation of that review
4. The title of review
5. The text of review
6. Total Likes/Dislikes count
7. Whether buyer is certified or not

The output achieved from the above, is the dataset of reviews and comments. The glimpse of the same is shown below:

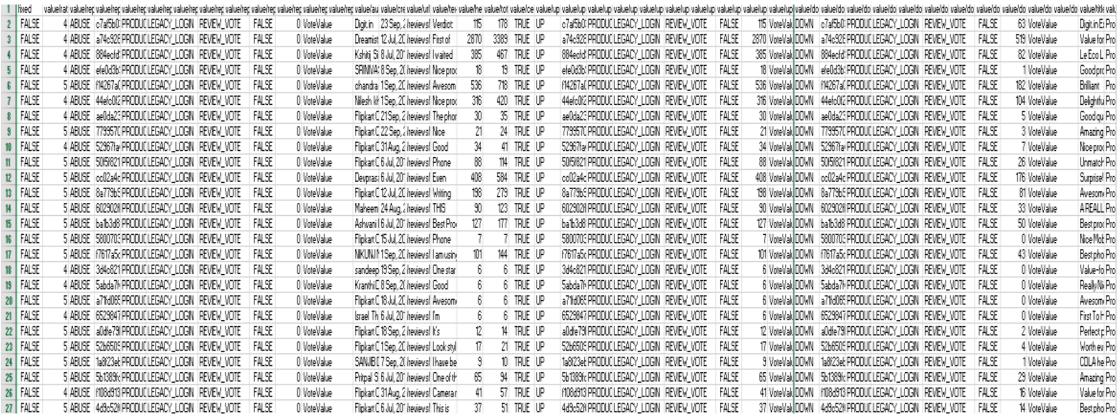
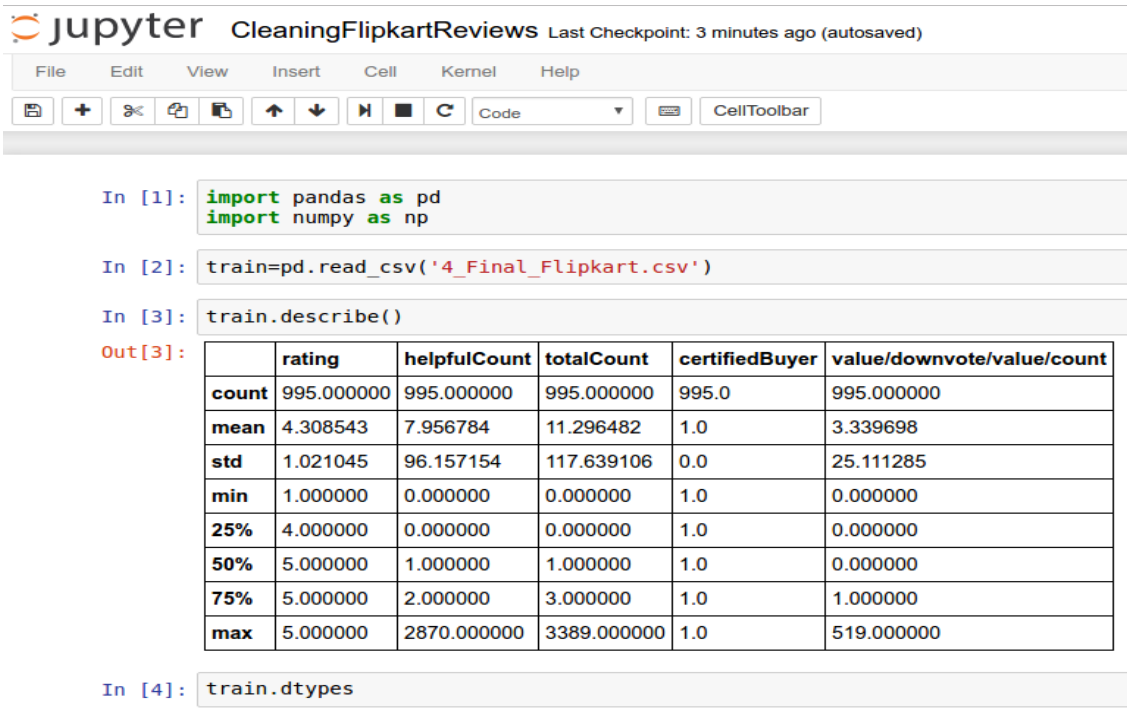


Figure 9. Shows ready to use data in a well-structured form.

2. *Data Cleansing Module:*

In this the various unnecessary keywords, null values etc. have been removed from the dataset. For data cleansing purpose, open source tool named "Panda" have been used [11]. Now we have a cleaned data, ready to be analyzed. The following image shows the data cleansing module of Panda.



	rating	helpfulCount	totalCount	certifiedBuyer	value/downvote/value/count
count	995.000000	995.000000	995.000000	995.0	995.000000
mean	4.308543	7.956784	11.296482	1.0	3.339698
std	1.021045	96.157154	117.639106	0.0	25.111285
min	1.000000	0.000000	0.000000	1.0	0.000000
25%	4.000000	0.000000	0.000000	1.0	0.000000
50%	5.000000	1.000000	1.000000	1.0	0.000000
75%	5.000000	2.000000	3.000000	1.0	1.000000
max	5.000000	2870.000000	3389.000000	1.0	519.000000

Figure 10. Data Cleansing Module.

Some results found after analysis:

1. No missing values
2. Increase in the number of reviews with the latest date.
3. All the buyers are certified, so can drop that column.(Not giving any information)
4. The rating column is preceded by apostrophe, so it is removed.

So the final dataset of 1005 observations having more than 10000 opinions is obtained.

rating	author	Date of Cr	text	helpfulCou	totalCount	certifiedBu	value/dow	Review title
4	Digit.in	23 Sep, 20	Verdict	115	178	1	63	Digit.in Expert Review For tho:
4	Dreamist	12 Jul, 201	First of all	2870	3389	1	519	Value for money
4	Kshitij Sing	8 Jul, 2016	I waited a	385	467	1	82	Le Eco Le 2 - You'll enjoy this c
4	SRINIVASA	8 Sep, 201	Nice produ	18	19	1	1	Good product with lack of hea
5	chandra ka	1 Sep, 201	Awesome	536	718	1	182	Brilliant
4	Nilesh kha	1 Sep, 201	Nice produ	316	420	1	104	Delightful
4	Flipkart Cu	21 Sep, 20	The phone	30	35	1	5	Good quality product
5	Flipkart Cu	22 Sep, 20	Nice	21	24	1	3	Amazing Cell phone...Value of
4	Flipkart Cu	31 Aug, 20	Good	34	41	1	7	Nice product
5	Flipkart Cu	6 Jul, 2016	Phone is	88	114	1	26	Unmatched
5	Devprasad	6 Jul, 2016	Even	408	584	1	176	Surprise
5	Flipkart Cu	12 Jul, 201	Writing	198	279	1	81	Awesome
5	Maheem J	24 Aug, 20	THIS	90	123	1	33	A REALLY GOOD BUDGET PHO
5	Ashvani Pa	6 Jul, 2016	Best Produ	127	177	1	50	Best product
5	Flipkart Cu	15 Jul, 201	Phone is	7	7	1	0	Nice Mobile
5	NIKUNJ M	1 Sep, 201	I am using	101	144	1	43	Best phone in this price segme

Figure 11. Final Opinion Dataset.

3. Direct Score Module:

In direct scoring a reviewer is supposed to give a rating from 1 to 5. Here the weightage of the rating and its classification can lie among any of the classes i.e. positive, negative or neutral [1]. For example if a customer gives a rating of less than 3, it is negative, if more than 3 it is considered to be positive and if it is exactly 3 it can be considered as neutral. So from this we have calculated a direct score based on these ratings. This has been scaled up to a scale of 5 in total. The final impact value from this score can be taken as -1 for negative, 0 for neutral or 1 for positive. Other methods that can also be applied here for deciding about the opinion of customer are Support Vector Machine (SVM) and Bayesian Classification etc. In direct score we will collect the review ratings provided by the customer. This is very influential parameter and carries closer concern about a customer's opinion about a product or service.

- Total no of rating counts: 1005
- Sum of the rating: 4322
- Average rating value: 4.3
- So the direct score is 4.3

4. Indirect Score Module

Here for having an accurate analysis it is important to note that considering only review rating is not always sufficient. But need to be aware about the reviewer's comments. For example if a reviewer is providing a rating of 4, from direct scoring we can easily say that he is positively commenting. But this necessarily not give the actual picture about the opinion of the reviewer. For example suppose a reviewer bought a mobile phone and as feedback he is giving a rating of 4 and the comment given by him is as following: "The battery does not last long". Here we have used corpus based approach for opinion classification. In this approach, In a large corpus of opinion related words, an existing pattern is derived with the support of seed list of opinion words. In this technique a list of seed opinion words are chosen. These are used to find the additional opinion words with conjunctions like AND etc. For example suppose our seed list has a word fast. This is sign of positive feedback. If in our dataset we have fast and sleek. It indicates clearly that sleek is also a positive word. Similarly rules for other conjunction words like BUT, NEITHER-NOR, EITHER-OR, OR etc. can also be used to extend the learning from the reviews. Finally a cluster has been developed for the positivity or negativity of opinion. From the cluster analysis a total score based on this is found. But there are challenges in this approach as well for example if some review says "I am trying to find a good mobile phone." Here good does not represent a positive indication and neither is it an negative indication. So this type of cases will have an impact on the accuracy of the score. So Indirect module is supposed to analyze the sentiments that are expressed indirectly or in a hidden way by a customer [2, 3, 4, and 5].

5. Integration Module

Though by now we have calculated a direct rating based score and an indirect review based score. But being a direct score we will use the J48 algorithm [8]. In this we have predicted the final score. And on the basis of this prediction has been made. For this purpose we have used WEKA 3.8. The results of same has been shown below in Part-B of the experiment section. It combines both direct and indirect scores and provide a trust score based on multiple dimensional feedback provided by a customer.
The formula used for integrating the direct and indirect score values is:

$$Score = (R+D+PI+V+C)/T*100$$

R: Rating value on the scale of 1-5
D: Date of creation of review on scale of 0-6.
PI: Positivity Index Value on scale of 0-10.
V: Vote Value on the scale 0-10.
C: Certified Buyer having value either 0 or 5.
T: Total Weighted having valuation as 36.

6. Overall Profile Builder.

Overall profile is to show the overall characteristic analysis of a seller. This graphical result collection, is to visualize the behavior of a seller.

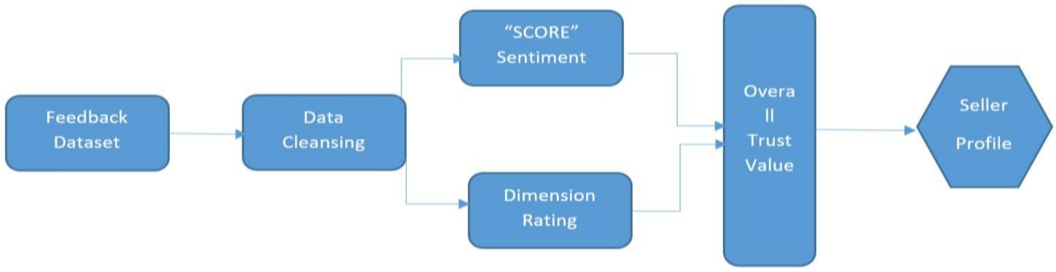


Figure 12. Customer Opinion Based Profile Building.

FINALIZING THE INTEGRATED SCORE

Now keeping the above discussed two phases of experimentation i.e.

1. The social media module where the social media dataset from the platform like "Twitter" is analyzed to get a score dependent on the same.
2. In second phase the product/seller analyses is based on the two aspects (i) On the direct reviews provided by the users/customers/visitors and the (ii) Dimension rating based on the comment statement analysis i.e. how customer is expressing on things like handling shipping quality of product, packaging or speed of customer care response etc.

Finally the above two can be combined to get a final impact on price of a product. This surely represent the global presence of internet that is reflected in the end price decision of price formation/prediction of any product or service. The figure below shows it:

CONCLUSION

By combining the social media and the customer reviews, model will be able to get a better insight about a customer. This gives an influence factor of the local conditions that he reads and build an opinion about a particular product. The different aspects achieved through this experiment lead to find about what is the nature of customer sentiment that exists and from what different areas these are being raised. This can be an influential factor about strategy on product selling either through direct selling or by online auctioning. Secondly it helps in deciding about the actual wave of thoughts, the reason for this is that large number of people participate in social media and review contribution. So this represent a real time picture of a product. Here the framework suggested will definitely aid in making impactful decisions about the product or service.

REFERENCES

[1] Kokate, Sushant, and Bharat Tidke. "Fake Review and Brand Spam Detection using J48 Classifier."

[2] Fang, Xing, and Justin Zhan. "Sentiment analysis using product review data." *Journal of Big Data* 2.1 (2015): 1.

[3] Peng, Qingxi, and Ming Zhong. "Detecting spam review through sentiment analysis." *Journal of Software* 9.8 (2014): 2065-2072.

[4] Pang, Bo, and Lillian Lee. "Opinion mining and sentiment analysis." *Foundations and trends in information retrieval* 2.1-2 (2008): 1-135.

[5] Liu, Bing. "Sentiment analysis and opinion mining." *Synthesis lectures on human language technologies* 5.1 (2012): 1-167.

[6] Smith, C. (2016, November 17). http://expandedramblings.com/. Retrieved from 170 Amazing Twitter Statistics and Facts: http://expandedramblings.com/index.php/march-2013-by-the-numbers-a-few-amazing-twitter-stats/

[7] SMITH, K. (2016, 5 17). *44 Twitter Statistics for 2016*. Retrieved from brandwatch.com: https://www.brandwatch.com/blog/44-twitter-stats-2016/

[8] Patil, Tina R., and S. S. Sherekar. "Performance analysis of Naive Bayes and J48 classification algorithm for data classification." *International Journal of Computer Science and Applications* 6.2 (2013): 256-261.

[9] Kumar, Sandeep, and Rahul Rishi. "Data collection and analytics strategies of social networking websites." *Green Computing and Internet of Things (ICGCIoT), 2015 International Conference on*. IEEE, 2015.

[10] Razi, Muhammad A., and Kuriakose Athappilly. "A comparative predictive analysis of neural networks (NNs), nonlinear regression and classification and regression tree (CART) models." *Expert Systems with Applications* 29.1 (2005): 65-74.

[11] McKinney, Wes. *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython*. "O'Reilly Media, Inc.", 2012.

[12] Krishnamurthy, Sandeep. "A Comparative Analysis of eBay and Amazon." *Intelligent Enterprises of the 21st Century. igi Global* (2004): 29-44.