International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 06 Issue: 06 | June 2019 www.irjet.net p-ISSN: 2395-0072

Opinion Mining from Customer Reviews for Predicting Competitors

Swathi Yadav[1], Dr. Deven Shah[2]

[1]Swathi Yadav, PG Student, Thakur College of Engineering and Technology, Mumbai, India
[2]Dr. Deven Shah, Professor, Dept. Information Technology, Thakur College of Engineering and Technology, Mumbai, India

---

Abstract – *Opinion mining is also known as sentiment analysis that is analyzing emotions. Emotions can be of two types positive and negative. Sentiment analysis or opinion mining is widely used in finding the emotions of the customer. Emotions of the customer can be analyzed by various means. Opinion mining is broadly connected to the voice of client materials, for example, reviews and overview reactions, on the web and various online platforms, and social insurance materials for applications that extend from promoting to client administration to clinical medication.*

In the current competitive business scenario, there is a need to analyze the competitive features and factors of an item that most affect its competitiveness. The evaluation of competitiveness always uses the customer opinions in terms of reviews, ratings and abundant source of information from the web and other sources. Providing an appropriate level of technique, will help the organization to know their competitors and will help them to improve where they lack.
In this project reviews are gathered from Amazon for a particular product and after that, reviews are uploaded and then sentiment analysis is performed on those reviews. In the second phase, reviews are classified into positive and negative. After classifying reviews pie-chart is generated, it shows the percentage of positive and negative reviews. In the fourth phase, the competitors are predicted, which helps us to know, the competitors of the particular product.

***Key Words*: Sentiment analysis, Machine learning, Opinion Mining, Web Mining, Pie- Chart.**

1.INTRODUCTION

Opinion mining is a kind of natural language processing for analyzing the mood of the public about a particular product. Opinion mining is a field of study that investigates people's opinion, feelings, sentiments or attitudes towards attributes like products, services, organizations, and occasions. Opinion mining is also known as sentiment analysis that is analyzing emotions.

Emotions can be of two types positive and negative. Sentiment analysis or opinion mining is widely used in finding the emotions of the customer. Emotions of the customer can be analyzed by various means. Opinion mining is broadly connected to the voice of client materials, for example, reviews and overview reactions, on the web and various online platforms, and social insurance materials for applications that extend from promoting to client administration to clinical medication. Sentiment analyses involve building a system to collect and categorize opinions about a product.

1.1 Benefits to the organization

Opinion mining is beneficial to an organization in several ways.

- It is very useful for marketers to evaluate the success of an ad campaign or new product launch.
- It can help the organization to determine which versions of a product or service are popular.
- It will also help to identify which demographics like or dislike particular product features.

1.2 Applications

Opinions are so important that whenever one needs to make a decision, one wants to know what others opinion is. This will be helpful for both individuals and organizations. So the procedure of sentiment mining has a wide extension for down to earth applications.

Singular buyers:

If an individual needs to buy an item, it is valuable to see a rundown of sentiments of different clients with the goal that he/she can settle on a superior choice. He/she will get to recognize what is the positive and negatives of an item before contributing cash on that item.

Associations and organizations:

Opinion mining is similarly, if not in any case progressively, imperative to business and organizations. For instance, it is basic for an item maker to know how buyers see its items and what the scope in the market to their item is, those of its rivals. This data isn't valuable for promoting and item benchmarking yet in addition helpful for item structure and item improvements.

1.3 Difficulties in Opinion Mining

When we talk about opinion mining, we should also highlight the difficulties that emerge in opinion mining. There are

several difficulties we should know, they are illustrated with the example below.

- A word that is considered to be certain in one circumstance might be viewed as negative in another circumstance.
  Example: Take the word "long" for example. Positive Opinion: If a client says "A P.C's battery life was long." Negative Opinion: If the client says "The laptop's start-up time was long". This dissimilarity implies that a sentiment framework prepared to gather sentiments of one type of product or product feature may not perform very well on another.
- Individuals don't in every case express an opinion in a similar way. Most traditional text handling depends on the fact that small dissimilarity between two bits of text does not change the meaning without a doubt.
  Example: In opinion mining, however, "the motion picture was extraordinary" is very different from "the motion picture was not extraordinary".
- Individuals can be conflicting in their statements. Most surveys will have both positive and negative reviews, which is somewhere manageable by breaking down each sentence one after another. Nonetheless, the more casual the medium (Twitter tweets or blog entries for model), the more probable individuals are to consolidate diverse suppositions in a similar sentence.
  Example: "The film bombed despite the fact that the lead on-screen actor rocked this time" is simple for us (human) to understand, but more troublesome for a computer to parse. Sometimes even some individuals face difficulty in understanding what someone thought based on a short piece of text because it lacks context, for example, "That motion picture was on a par his last one" is totally subject to how the person is expressing the opinion thought of the last movie.

2. LITERATURE REVIEW

Table -1: Literature Review table

Sr. No	Paper Title	Authors	Year	Description	Gap
1	Mining Competitors from Large Unstructured Datasets.	George Valkanas, Theodoros Lappas, and Dimitrios Gunopulos	2017	A formal definitions of competitiveness between two items, which we validated both quantitatively and qualitatively.	It is domain specific, that is considered only for restaurants. Only three types of competitors are analyzed top, middle and bottom.
2	Identifying comparative customer requirements from product online reviews for Competitor Analysis	Jin, Jian, Ping Ji, and Rui Gu	2016	How to select a small number of opinionate sentences from product online reviews for competitor analysis is investigated.	Results should be visualized in an interactive graphical user interface. How to compare products with the help of big opinionate product reviews
3	Mining competitor relationships from onlinenews: A network-basedapproach	Z. Ma, G. Pant, and O. R. L. Sheng	2011	It proposes and evaluates an approach that exploits company citations in online news to create an intercompany network whose structural attributes are used to infer competitor relationships between companies.	Do not examine news stories written in another language. Do not predict future competitor relationships.
4	Estimating aggregate consumer Preferences from	R. Decker and M. Trusov	2010	An econometric framework that can be applied to turn the plentitude of individual	Fake reviews are not detected. Time-consuming

	online product reviews			consumer opinions made available by online product reviews into aggregate consumer.	
5	Identifying customer preferences about tourism products using an aspect based opinion mining approach.	E. Marrese-Taylor, J. D. Vel´asque z, F. Bravo- Marquez, and Y. Matsuo	2013	Tourism product reviews available on web sites contain valuable information about customer preferences that can be extracted using an aspect-based opinion mining approach	Not domain sensitive: Specific sentences regarding context or domain dependent topics need to be specially treated.

3. SYSTEM FLOW

Phase.1
Extraction of reviews:

This is the first phase of the project, in this phase, the reviews are extracted from an Amazon so that the reviews would be useful to predict the competitors.

Product reviews exist in enormous forms, online: websites committed to a particular kind of product (such as MP3 player or movie pages), destinations for papers and magazines that may feature reviews (like Rolling Stone or Consumer Reports), websites that couple reviews with business (like Amazon), and websites that have some expertise in gathering proficient or customer reviews in an assortment of zones. Less formal reviews are accessible on dialog sheets and mailing list files, just as in Usenet by means of Google Groups. Clients additionally remark on items in their own sites and websites, which are then amassed by sites such as Blogstreet.com, AllConsuming.net, and onfocus.com. When endeavoring to find data on an item, a general web seek turns up a few valuable locales, yet getting a general feeling of these surveys can be overwhelming or tedious.
In this domain, the reviews are extracted from Amazon and the dataset is prepared in the excel sheet and then uploaded to the system.

Phase.2
Sentiment analyses:

This is the second phase of the project, in this phase, the extracted reviews are analyzed and sentiments are detected. And then the reviews are classified into positive and negative.

Sentiment Analysis examines the problem of studying reviews uploaded by customers on various social platforms online, about forums, and gadgets businesses, regarding the opinions they have about a product, service, event, person or idea.

Use of Sentiment Analysis is to classify a text to sentiments. Sentiments can be classified into binary (positive or negative) or multi-class (3 or more classes) problem, depending on the dataset and the reason.

You can find the same or different reviews, depending on the customer's perspective. These reviews can then be classified, through sentiment analysis.

To classify sentiments into positive or negative, first, we have to perform pre-processing. POS-tagging is must be carried on to data in order:

- To reduce the noise of the text.
- To reduce the dimensionality
- Assist in the improvement of classification effectiveness.

POS-Tagging

POS tagging is nothing but Parts of Speech Tagging. Parts of speech (POS) are explicit lexical classes to which words are relegated, in light of their syntactic context and job. As a rule, words can be categorized into one of the accompanying real classes.

- **N(oun):** This normally means words that portray some article or substance, which might live or nonliving. A few models would be the fox, hound, book, etc. The POS label image for things is N.
- **V(erb):** Verbs are words that are utilized to depict certain activities, states, or events. There is a wide assortment of further subcategories, for example, assistant, reflexive, and transitive action words (and some more). Some ordinary instances of action words would run, bouncing, read, and compose. The POS label image for action words is V.
- **Adj(ective):** Adjectives are words used to depict or qualify different words, ordinarily things and thing phrases. The expression wonderful blossom has the thing (N) bloom which is portrayed or qualified to utilize the descriptor (ADJ) delightful. The POS label image for descriptive words is ADJ.
- **Adv(erb):** Adverbs generally go about as modifiers for different words including things, descriptive words,

action words, or different intensifiers. The expression excellent blossom has the verb modifier (ADV) very, which alters the descriptor (ADJ) lovely, showing how much the bloom is delightful. The POS label image for verb modifiers is ADV.

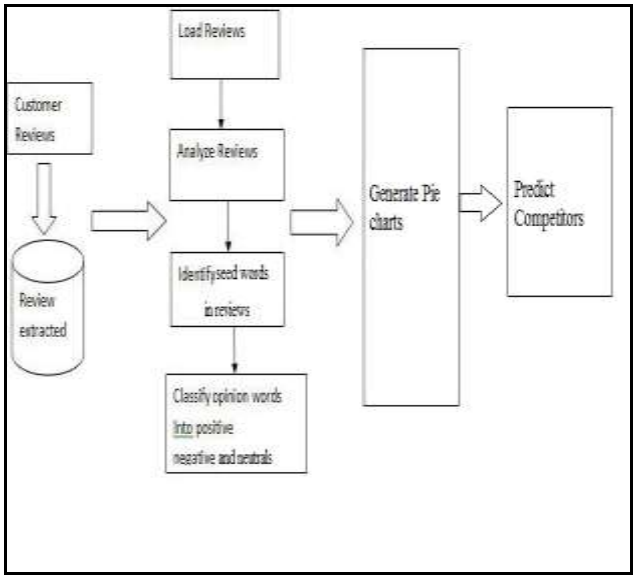

Figure -1: System Flow

Phase.3

Generate pie-charts: This is the third phase of the project, at this phase classified result will be represented using graphs. It shows the percentage of positive, negative words in the document. It helps us to make a decision on whether the sentence implies a positive or negative impact on a product to make a better decision.

The system presents the users the pie-charts, which will show the positive and negative sentiments based on sentiment analysis. Hence, in this study we find out that sentiments play a very important role in knowing competitors, we consider all the possible features of the particular product and according to it, positive and negative sentiments are classified.

Phase.4

Predict Competitors: This is the fourth phase of the project. The main focus was to use the reviews and comments obtained from the users about an online product, for predicting the competitors of that product or item.

In this phase, the scores are given to the product according to their aspects or features. Then these products are ranked accordingly with their positive and negative sentiments so that we can judge or determine the competitors of the particular product. And even we can know, which product is better or week for the particular aspects, and which else are the competitors.

4 RESULTS AND DISCUSSIONS:

The proposed architecture is a help to predict the competitors of the particular product. Below are some snapshots to which will help to understand the concept.

Figure 2: Snapshot of Uploading Dataset

Figure 2 is the snapshot of uploading Dataset. The dataset is made in the form of excel sheet namely final_main.xls. This file contains the reviews, which are in a structural format. The reviews, which is collected from Amazon Website is prepared in a structural format so that it would be easy to upload.

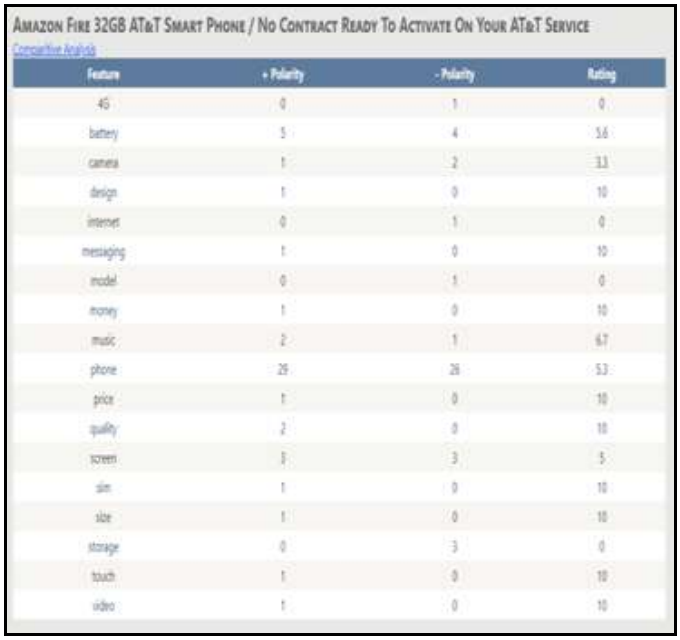

Figure 3: Sentiment analysis based on features

The above Figure 3 shows the sentiments analysis result. It shows the polarities of each feature of the particular product. Feature by classification of a product is very helpful to know the competitors of that product.

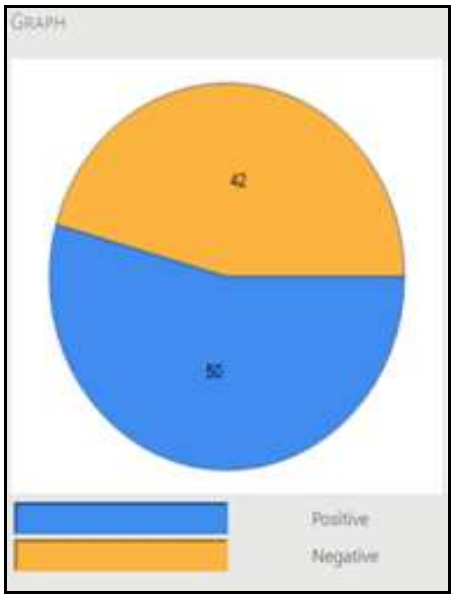

Figure 4: Snapshot of pie-chart of a positive and negative sentiment of the product.

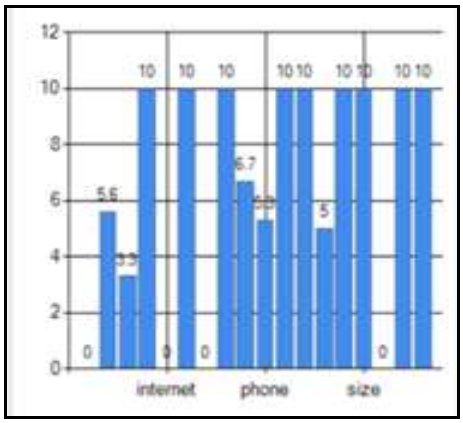

Figure 5: Snapshot of graph for ratings of the product

The above Figure 4 and Figure 5 shows, the positive, negative polarities and the ratings of the product. These polarities and, the ratings are calculated by sentiment analysis.

In the above two figures, these ratings and polarities are represented in the form of graph and pie-chart. This helps us to know the positivity and the negativity of the particular product, and also we can find the weakness of the product.

Brand_Name	Positive	Negative
Acer	80	78
Amazon	50	42
Apple	162	126
Cedar Tree Technologies	20	16
Digital SNITCH	50	42
HTM	15	12
Huawei	45	33
Jiuhe	36	36
Lenovo	43	33
Nokia	15	12
Samsung	14	11
Ulefone	34	30
VKworld	55	39

Figure 6: Snapshot of Comparative analysis

The above Figure 6, shows the snapshot of comparative analysis. This helps us to predict the competitors of a particular product. All the brand names are listed with their positive and negative polarities, which would assist us to predict the competitors.

5. CONCLUSION

The main aim of this study was to use the reviews and comments obtained from users about an online product, for predicting the competitors of that product or item. This study has led us to devise the way so that, instead of going through each comment one by one, a user can get the feedback of all users in graphical format. It will help the user to know the good and bad quality of the product. In this study, competitors are analyzed based on aspects of the products specified by the users. They can even know where their product lack from others. The user can see the competitors of the particular product, classified in terms of positive and negative reviews. In the future, the work can be extended by classifying reviews in more details and investigate a different kind of feature to make more accurate predictions.

ACKNOWLEDGMENT

I would like to take the opportunity to express my sincere thanks to my Guide **Dr. Deven Shah**, Vice Principal, Professor, IT Department, TCET and ME-IT Coordinator **Dr. Bijith Marakarakandy**, Professor, IT Department, TCET for his invaluable support and guidance throughout my P.G. research work. Without his kind guidance & support, this was not possible. I am grateful to him for timely feedback which helped me track and schedule the process effectively. His time, ideas and encouragement that he gave helped me to complete my project efficiently.

I would like to thank **Dr. Kamal Shah**, Dean, Professor, IT Department for her guidance and encouragement throughout my Post Graduation.

I would also like to thank **Dr. B. K. Mishra**, Principal, Thakur College of Engineering and Technology, for his encouragement and for providing an outstanding academic environment, also for providing adequate facilities. I am thankful to all my M.E. teachers for providing advice and valuable guidance.

I also extend my sincere thanks to all the faculty members and the non-teaching staff and friends for their cooperation. Last but not least, I am thankful to all my family members whose constant support and encouragement in every aspect helped me to complete my project.

REFERENCES

[1] Edison Marrese-Taylora, J. D.-M. (2013). Identifying Customer Preferences about Tourism Products using an Aspect-Based Opinion Mining Approach. 17th International Conference in Knowledge Based and Intelligent Information and Engineering Systems - KES2013 , 182-191.

[2] George Valkanas, T. L. (2017). Mining Competitors from Large Unstructured Datasets. IEEE Transactions on Knowledge and Data Engineering , 1-18.

[3] Zhongming Maa, ⇑. G. (2011). Mining competitor relationships from online news: A network-based approach. Electronic Commerce Research and Applications , 418-427.

[4] BIBLIOGRAPHY Reinhold Decker, Michael Trusov. (2010). Estimating aggregate consumer preferences from online product reviews. Intern. J. of Research in Marketing 27, 293–307.

[5] Ping JI,Jian JIN,. (2015). Extraction of Comparative Opinionate Sentences from Product Online Reviews. 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD) , 1777-1785.

[6] Kaiquan Xu, Stephen Shaoyi Liao, Jiexun Li, Yuxia Song (2011). Mining comparative opinions from customer reviews for Competitive Intelligence. Decision Support Systems, 743–754.

[7] BIBLIOGRAPHY Mr. Vinayak Hegde1, Ms. Karthika P 2, Ms. Madhu M G (2015). Opinion Mining And Market Analysis. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 10, Number 10, pp. 25629-25636.

[8] Larissa A. de Freitas, Renata Vieira, "Ontology-based Feature Level Opinion Mining for Portuguese Reviews", ACM- 978-1-4503-2038- 2/13/05 WWW 2013

[9] E. Cambria, B. Schuller, Y. Xia, and C. Havasi, "New avenues in opinion mining and sentiment analysis," IEEE Intelligent Systems, vol. 28, no. 2, pp. 15–21, 2013.

[10] Padmapani P. Tribhuvan, S.G. Bhirud, Amrapali P. Tribhuvan (2014). A Peer Review of Feature Based Opinion Mining and Summarization. I nternational Journal of Computer Science and Information Technologies, Vol. 5 (1), 247-250.

[11] Li Chen, Luole Qi, Feng Wang (2012). Comparison of feature-level learning methods for mining online consumer reviews. Expert Systems with Applications 39, 9588–9601.

[12] Yi Fangy, Luo Siy, Naveen Somasundaramy, Zhengtao Yuz (2012). Mining Contrastive Opinions on Political Texts using Cross-Perspective Topic Model. WSDM'12, 8–12.

[13] Yu Peng, Raymond Chi-Wing Wong and Qian Wan (2007). Finding Top-k Preferable Products. JOURNAL OF LATEX CLASS FILES, VOL. 6, NO. 1,, 1-20.

[14] Amruta SankheÅ, Prachi GharpureÅ (2014). Feature Based Sentiment Analysis for Online Reviews in Car Domain. International Journal of Current Engineering and Technology, E-ISSN 2277 – 4106, P-ISSN 2347 - 5161

[15] Yu Peng, Raymond Chi-Wing Wong, and Qian Wan (2012). Feature Based Sentiment Analysis for Online Reviews in Car Domain. IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING, VOL. 24, NO. 10, 1774-1788.

[16] T. Yano, W. Cohen, and N. Smith. Predicting response to political blog posts with topic models. In NAACL/HLT, pages 477{485, 2009}.

[17] J. Yi, T. Nasukawa, R. Bunescu, and W. Niblack. Sentiment analyzer: Extracting sentiments about a given topic using natural language processing techniques. In ICDM, pages 427{434. IEEE, 2003.

[18] C. Zhai, A. Velivelli, and B. Yu. A cross-collection mixture model for comparative text mining. In SIGKDD, pages 743{748, 2004.

[19] M. Zhang and X. Ye. A generation model to unify topic relevance and lexicon-based sentiment for opinion retrieval. In SIGIR, pages 411{418. ACM, 2008.

[20] W. Zhang, C. Yu, and W. Meng. Opinion retrieval from blogs.In CIKM, pages 831{840. ACM, 2007.