International Journal of Computer Sciences and Engineering Open Access
Survey Paper Vol.8, Issue.8, August 2020 E-ISSN: 2347-2693

Survey on Sentimental Analysis and Visualization of Reviews

Bhanu Chugh[1*], Sejal Arya[2], Mayank Pandita[3], Tanmay Jain[4], G.V. Bhole[5]

[1,2,3,4,5]Dept. Of Information Technology, Bharati Vidyapeeth(Deemed to be University) College of Engineering, Pune, Maharashtra, India

[*]*Corresponding Author: bhanuchug@gmail.com, Tel.: +91-7387516869*

DOI: https://doi.org/10.26438/ijcse/v8i8.3033 | Available online at: www.ijcseonline.org

Received: 29/Jul/2020, Accepted: 17/Aug/2020, Published: 31/Aug/2020

***Abstract*—** The time when the food ordering websites are filled with tons of review data regarding the quality and quantity of food. One can extract tons of conclusions while analyzing it. In this survey paper, we will study various techniques to analyze the data including algorithms like Naive Baye's, SVM, etc. and their outcomes in the field of data mining and sentimental analysis. Sentimental analysis is a boon for the restaurant owners as they can restructure their unique selling points and services. Customers can indeed use the data to filter the restaurants according to the area, cuisines, dining time, etc. to make an opinion. Also, in the end, polarising the high-quality dataset into positive and negative vows to visualize it for the customers. Tableau the most majestic tool can help us to do the same. Working on emoticons as well as text will take us to a hard way to complete the study.

***Keywords*—** Opinion Mining, Sentimental Analysis, SVM, Tableau.

I. INTRODUCTION

In today's world where everyone wants to comment on every process, function, or food being served. Social media website is also a boon for data analysts. From Facebook, Twitter to e-commerce websites like Amazon, Flipkart is flooded with a huge amount of customer feedback[1].Different algorithms are already in the market to judge the sentiments of customers[2]. Sentimental analysis of the comments and reviews on different websites by lakhs of customers is the need of the hour. Review data is not only available easily but also the internet is flushed with bytes of data every second. Understanding the given data and judging it based on several aspects sums up Opinion Mining. When aspects are either positive or negative then it is termed as Sentimental analysis[5]. This type of analysis can help the government to take certain actions in the cyber world [6]. The trend of Sentimental analysis started from the 20th century and now has been the most trending among the researchers.

Apart from that, we can use this analysis to analyze a certain amount of restaurants and different cuisine food being served there.

Our main objective is to focus on :
(i) What type of food does the customer want?
(ii) Is the customer satisfied with the food?
(iii) Is communication possible between the user and a restaurant?[10]

In this paper, we will first study a bit of research being done in the field of sentimental analysis, opinion mining, and data visualization which suffice section II, then coming up with a methodology and conclusion to solve the problem statement in section III and finally identifying its scope for future researches in the last section.

II. LITERATURE SURVEY

In paper the by Santhosh Kumar, a survey study on sentimental analysis and opinion mining on text from amazon based mobile phone reviews was made using 3 basic algorithms Naive Bayes, Logistic Regression, and SentiWordNet.The study was done on 3 different aspects namely Recall, Precision, and F-Measure for each product review respectively for each algorithm. From the overall results, it was clear that Naive Bayes came out to be the best algorithm while mining or studying the reviews. They also proposed a model which can help to mine the future reviews from websites like Snapdeal, Flipkart, etc. which are e-commerce websites. But we will try to extend our research to a food delivering website such as Zomato, Swiggy, etc. [1].

Vipin Kaur worked with two types of algorithms while dealing with sentiments of data from the e-commerce website Amazon. She classified analysis into supervised and unsupervised approaches consisting of 3 types of algorithms namely, Naive Bayes, Support Vector Machine(SVM, which is a machine learning algorithm), and SO-PMI-IR technique. The results came out to be the supervised structure gave astonishing results for one-lined reviews while unsupervised structure responded adequately to the big piece of data. The aspects were accuracy, precision, etc. and concluded in favor of the choice of dataset[2].

Robert in 2009, surveyed over thousands of Biomedical Research Papers and finally rested a fact that NLP-NG is the future. Here, NLP-NG includes 3 parts from language processing to the visualization of data being entered. He strongly focused on the fact of Normalisation of the dataset which can eventually help in maintaining a high order database. Also, he wanted to implement the same to the various Chat boxes like Google, etc. A large scale normalization can be implemented while refining the datasets[3].

One year ago in paper, Rakibul created an NLP platform for analyzing Twitter data. They focused on the intensively increasing twitter data and tried to measure its polarity. They also used the BoW and TF-IDF model to work on the polarity of the sentences being tweeted. Apart from polarity, they were able to provide amazing insights for the user. One of their aspects was accuracy which made them succeed in this path. Their classifier can be used as a sentimental analysis tool for any of the datasets further[4].

Way back in 2017, Perera made aspect-based mining for the restaurant reviews. Different extraction tools like SentiWordNet, POS tagging, and dependency parser were used to analyze the food reviews. They concluded that their search scarce various other keywords to be analyzed and emoticons on the respective restaurants. Though they tried to give a negative, positive, or neutral polarity at the end. We can propose the same model which can work with emoticons as they now constitute a major part of reviewing[5].

Sudeep, tried to analyze the twitter data using StanfordNLP libraries and SaaS. Also, they used twitter4j to extract real-time data from twitter and make the analysis. The final analysis was given polarity of either negative or positive to every review. Their future scope refers to help the government in analyzing the sentiment of reviews and work on their cyber policies and protect social media websites[6].

Pooja and Arjit did a sentimental analysis of reviews from different e-commerce websites and tried to polarize the same using a calculable meter. The used Google Chart API to visualize their results. The categories were price, size, speaker and screen, etc. as for the mobile phone reviews. They aimed to answer the queries being asked by the user of every aspect. Future scope tends to create a system that takes and analyses the query of users automatically[7].

In the paper , Wang used the Spline algorithm to visualize the analysis using the lower descent of the limit approximation. The earlier approximation visualized using the same algorithm but he tried to reduce the distance between two values to furnish the correct result. Knots and controlled vertices overall helped him to achieve a visualized approach for the same[8]. While in other paper , the visualization is done using Tableau. It is a magnificent tool that can be easily used to visualize. Moreover, they used Hadoop as a platform to analyze the data and two of its types were Apache Pig and Apache Hive. In contrast, Pig was easier to understand and connect to the historical databases. Further, their research can be taken to MySQL, R, and Python to visualize the data[9].

In paper by Aljoharah used R as a tool to sentimentally analyze the reviews on various books on Amazon. The reviews are classified based on whether neutral, positive, or negative. These deviations were visualized using various charts such as Bar Graph, Word Cloud, and Packed Bubbles. A final dashboard shows all the deviations on a single screen. Emoticons are again not used as sentimental raw data which again is a drawback for the proper mining of opinion of reviews[10].

In other paper, the sentimental analysis of reviews posted on Amazon is done. The paper is made to prove that the spam reviews are not posted on Amazon as it's a part of its algorithm checking process. One cannot post anything spam as defined by some other researchers. Also, they did a contrasting search between different types of algorithms including Naive Bayes, SVM, etc. Polarity is also being extracted from the same[11].

But according to Mika, researches in the field of sentimental analysis have been increased eventually since the end of World War II, earlier the searches were based on communalism, socialness, etc. but now after 2004, 99% increase is seen in the number of sentiment researches. Nowadays, these researchers have moved forward to social media handles like Twitter and Facebook. From elections to medicine for COVID-19 we have never-ending real-time reviews on social media handles. Their say is " Pen is mightier then sword" proves that there can be a 50 fold increase of research papers in this context. Also, the sentimental analysis will one day become a standardized procedure for every social media website [12].

Table 1

S.No.	Authors Name	Paper Name	Method Used	Description
1.	Aljoharah Almjawel, Sahar Bayoumi, Dalal Alshehri, Soroor Alzahrani, Munirah Alotaibi	Sentiment Analysis and Visualization of Amazon Book's Reviews	R and Tableau	Visualizing the sentimentally analyzed customer comments on a dashboard.
2.	Vipin Deep Kaur	Sentimental Analysis of Books Reviews using Unsupervised Semantic	Naive Bayes, Support Vector Machine(SVM), and SO-PMI-IR	Gave a glimpse of contrast between supervised and unsupervised approaches.

		Orientation and Supervised Machine Learning Approaches		
3.	Santhosh Kumar K L, Jayanti Desai, Jharna Majumdar	Opinion Mining and Sentiment Analysis on Online Customer Review	Naive Bayes, Logistic Regression, and SentiWordNet	Study on different aspects Recall, Precision, and F-Measure
4.	Robert P. Futrelle, Jeff Satterley, Tim McCormack	NLP-NG- A new NLP System for Biomedical Text Analysis	NG-CORE, NG-DB, and NG-SEE	Normalization of the dataset to maintain the high quality of research
5.	I.K.C.U. Perera, H.A. Caldera	Aspect Based Opinion Mining on Restaurant Reviews	SentiWordNet, POS tagging, and dependency parser	Neutral, Positive, and Negative polarity to sentiments.

III. METHODOLOGY & CONCLUSION

After, referring a lot of researches and understanding the current scenario of sentimental analysis and opinion mining. We finally conclude to the model of analyzing and normalizing the raw data. Fetching the desired sentiments by assigning them certain polarity data. Naive Bayes or R can be used to extract the same. After the scores are recorded then mending it into a data table and visualize the same using Tableau.

From the researches above, we also conclude that the Naive Bayes algorithm came out to be the best-suited algorithm for the Sentiment analysis of comments. Further, we will work on a random forest algorithm to structure the clustered data[13].

IV. FUTURE SCOPE

The authors in the future are thinking to build a system that can help customers and restaurant owners to understand customer sentiments. Customers can easily select and can be deflected to restaurants having high sentiment reviews in specific areas. Comparison among the scope like dining hours, parking, Wi-Fi, cuisine served can help in prioritizing the customer's demands. Further, restaurant owners can work on improving their Cuisine demand and food taste to attract customers. Also, cluster of data can be refined using various data mining techniques of the restaurant data and further can help in proper setting of visualizable and useful data [14]. Various data verifying and validating techniques can be used to extract the true data.

REFERENCES

[1] Kumar, KL Santhosh, Jayanti Desai, and Jharna Majumdar. "*Opinion mining and sentiment analysis on online customer review*." In *2016* IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), pp. **1-4**. IEEE, **2016**.

[2] Kaur, Vipin Deep. "*sentimental analysis of Book Reviews using Unsupervised Semantic Orientation and Supervised Machine Learning Approaches*." In Second International Conference on Green Computing and Internet of Things (ICGCIoT). **2018**.

[3] Futrelle, Robert P., Jeff Satterley, and Tim McCormack. "*NLP-NG—A new NLP system for biomedical text analysis*." In **2009** IEEE International Conference on Bioinformatics and Biomedicine Workshop, pp. **296-301**. IEEE, **2009**.

[4] Hasan, Md Rakibul, Maisha Maliha, and M. Arifuzzaman. "*Sentiment Analysis with NLP on Twitter Data.*" In 2019 International Conference on Computer, Communication, Chemical, Materials and Electronic Engineering (IC4ME2), pp. **1-4**. IEEE, **2019**.

[5] Perera, I. K. C. U., and H. A. Caldera. "*Aspect based opinion mining on restaurant reviews.*" In 2017 2nd IEEE International Conference on Computational Intelligence and Applications (ICCIA), pp. **542-546**. IEEE, **2017**.

[6] Kisan, Hase Sudeep, Hase Anand Kisan, and Aher Priyanka Suresh. "*Collective intelligence & sentimental analysis of twitter data by using StandfordNLP libraries with software as a service (SaaS)*." In 2016 IEEE international conference on computational intelligence and computing research (ICCIC), pp. **1-4**. IEEE, **2016**.

[7] Kherwa, Pooja, Arjit Sachdeva, Dhruv Mahajan, Nishtha Pande, and Prashast Kumar Singh. "*An approach towards comprehensive sentimental data analysis and opinion mining*." In 2014 IEEE International Advance Computing Conference (IACC), pp. **606-612**. IEEE, **2014**.

[8] Wang, Shin-Ywan, and Leonard A. Ferrari. "*Automatic Data Visulization Using Spline Functions*." In 1990 Conference Record Twenty-Fourth Asilomar Conference on Signals, Systems and Computers, 1990., vol. **1**, pp. **413-413**. IEEE Computer Society, **1990**.

[9] Bhardwaj, Priti, and Niyati Baliyan. "*Hadoop based Analysis and Visualization of Diabetes Data through Tableau*." In 2019 Twelfth International Conference on Contemporary Computing (IC3), pp. **1-5**. IEEE, **2019**.

[10] Almjawel, Aljoharah, Sahar Bayoumi, Dalal Alshehri, Soroor Alzahrani, and Munirah Alotaibi. "*Sentiment Analysis and Visualization of Amazon Books' Reviews*." In 2019 2nd International Conference on Computer Applications & Information Security (ICCAIS), pp. **1-6**. IEEE, **2019**.

[11] Fang, Xing, and Justin Zhan. "*Sentiment analysis using product review data*." Journal of Big Data 2, no. 1 (2015): 5.

[12] Viking Mäntylä, Mika, Daniel Graziotin, and Miikka Kuutila. "*The Evolution of Sentiment Analysis-A Review of Research Topics, Venues, and Top Cited Papers*." arXiv (2016): arXiv-**1612**.

[13] Fernandes, Marie. "*Data Mining: A Comparative Study of its Various Techniques and its Process*." International Journal of Scientific Research in Computer Science and Engineering 5, no. 1 (2017): **19-23**.

[14] R.S. Walse, G.D. Kurundkar, P. U. Bhalchandra, "*A Review: Design and Development of Novel Techniques for Clustering and Classification of Data*," International Journal of Scientific Research in Computer Science and Engineering, Vol.**06**, Issue.**01**, pp.**19-22**, **2018**

AUTHORS PROFILE

Mr. Bhanu Chugh is pursuing a Bachelor of Technology in Information Technology Department at the Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune. His research interests include visualization using Tableau, Sentimental analysis using Rstudio, Python, and Web-development.

Ms. Sejal Arya is pursuing a Bachelor of Technology in Information Technology Department at the Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune. Her research interests include visualization using Tableau.

Mr. Mayank Pandita is pursuing a Bachelor of Technology in Information Technology Department at the Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune. His research interests include Sentimental analysis using Rstudio.

Mr. Tanmay Jain is pursuing a Bachelor of Technology in Information Technology Department at the Bharati Vidyapeeth (Deemed to be University) College of Engineering, Pune. His research interests include visualization using Tableau.

Prof. G.V. Bhole is currently working as Assistant Professor in the Department of Information Technology, Bharati Vidyapeeth (Deemed to be University), College of Engineering, Pune-43 India. He has published more than 10 research papers in reputed international journals including Scopus Journal and conferences which are available online. His main research work focuses on Web Development, Software Engineering, Data Science, and opinion mining.

