



Sentimental Analysis and Visualization of Food Reviews from Zomato using Tableau

Bhanu Chugh¹, Mayank Pandita², Sejal Arya³, Tanmay Jain⁴, G.V.Bhole⁵

¹Dept. Of Information Technology, Bharati Vidyapeeth(Deemed to be University) College of Engineering, Pune, Maharashtra, India : bhanuchug@gmail.com

²Dept. Of Information Technology, Bharati Vidyapeeth(Deemed to be University) College of Engineering, Pune, Maharashtra, India : mnkpandita26@gmail.com

³Dept. Of Information Technology, Bharati Vidyapeeth(Deemed to be University) College of Engineering, Pune, Maharashtra, India : sejalarya18@gmail.com

⁴Dept. Of Information Technology, Bharati Vidyapeeth(Deemed to be University) College of Engineering, Pune, Maharashtra, India : jain15tanmay1508@gmail.com

⁵Dept. Of Information Technology, Bharati Vidyapeeth(Deemed to be University) College of Engineering, Pune, Maharashtra, India : gvhole@bvucop.edu.in

ABSTRACT

In today's digital era where everybody has their own opinion and so can post various feedbacks regarding different products. Reviews are critical in nature and can affect the market value of a product. Amazon, Zomato, Quora, Twitter, etc . are few such platforms that provide space for product reviews. Since reviews are textual and diversified; a collective representation will help users to summarize their opinions. Honest food reviews and its analysis is still a challenge. Sentiment Analysis or opinion mining indicates the utilization of computational linguistics, textual analysis, biostatistics, and NLP (Natural Language Processing) to consistently recognize, skin, compute and grasp effective levels and abstract information. In this paper, we used the database of 'Zomato Pune' downloaded from Kaggle which is used as a basic dataset to summarize the food reviews by customers into 3 basic categories whether positive, negative or neutral based on the food being served at the restaurants. A visual analytical tool is being used linked with Tableau to give clear and interactive charts: Bar Chart, Word-Cloud, and Packed Bubble as visualization techniques. Charts will help users to compare multiple options interactively at the same time.

Key words: Food Reviews, NLP, Opinion Analysis, Sentiment, Tableau, Text Visualization.

1. INTRODUCTION

In a globally networked world, where the internet is a boon to the immediate feedback of sentiments that are based on emotions. People review the services and products being offered to them on various websites. These websites not only help customers to share their reviews but also allows other

customers and business owners to improve the range of their services. Users with different backgrounds furnish their reviews in different languages and scripts which are gold for the opinion miners. Online visualization of the user's feedback is necessary, which is indeed responsible for the overall insight of a user and is increasing the indulgence of individuals in the field of opinion mining and analysis. [1] Sentiment analysis is responsible for the wrong beliefs prevailing in the user's mind regarding any service or product. It chooses the best of comments and analyzes the same for a specific set of interests.[2]

For e.g.: People buying products on Amazon look for quality and price efficiency and decide the fate of the product and company indeed. [3]

Our aim is to propose a visualized model of the customer feedback on various restaurants, to the users and restaurants regarding the quality of food served to the customers. Enhancing Restaurants to deliver honest advertisements and improve product quality. To reach the desired aim, a sentimental analysis approach system of twitter in the tableau is used to judge the set of textual reviews but here we used a large food dataset.

One can tackle the following hurdles from the given system: (i) which restaurant is getting the highest positive on plate reviews? (ii) Can a user find top-rated restaurants in a specific locality? (iii) Comparing several restaurants according to opinion reviews?

Paper is branched into the following sections: section II showing a literature survey of old researches on sentiment analysis and data analytics related to the aim; section III shows the furnished visualized platform; section IV frames the result of the unified system with the exclusive dataset from Zomato and section V synopsis of the conclusion and section VI refers to the future work for the problem statement.

2. LITERATURE REVIEW

A search in early 2017 has been done over all the research to make a count of the uncovered areas for the ongoing literature sections by Godwin. The dataset was gathered from multiple fields like Scopus, Web of Science databases, Education Full Text, ERIC, and Engineering Village. Subsequently, from linguistic analysis, a visualization technique was developed termed as a co-occurrence network. This approach granted each cluster in the network as a different node shape and color. Lastly, the co-occurrence network-assisted researchers to focus on the important breach, and accordingly showcased swift and trending tricks to enrich further analysis related to the areas that haven't been explored.[4]

Similar reviews were also furnished under a survey by Aljammal and Salameh regarding the visualization of the software tools. Their main agenda focused to inspect the Software Evolution Visualization's main target, analyze the terminology which was served to the respective toolkit, and investigate the key source of facts that visualized the evolution of new software techniques. Various techniques classifying the software evolution are metaphor-based, matrix-form, notation-form, graph-form, etc. Notation form was ignored as compared to graph-form which in turn was more favored. The study was responsible for achieving a standardized form for assisting researchers, conservator developers, to acquire accurate data responsible for visualized and evolved software to an entity.[5]

Ontology visualization prescribed by Mikhailov stated the systematic reviews of the literature in the same operational field. Many papers regarding the visualization using ontology were read and tested. There were more than 50 papers from 2006 which were examined. After that, the papers were analyzed to make a decision about which visualization techniques were to be executed in ontology visualization. It was established that the regularly used visualization techniques were: Concept diagrams, Radial tree, 2-D and 3-D graphs, and other concepts. The most frequently used technique was visualization using 2-D plain graphs.[6]

In paper [7], the authors suggested a collective structure to anticipate the purpose of helping customers to choose the best of given choices. It involved a lot of informative data from discrete webpages. The structure focused to analyze each and every word from the feedback to describe it's the orientation and which resulted in separating nouns as marked functions and opinions were adjectives. Then, the functions extracted from the feedback were judged as are they neutral align, positive align, or negative align. Finally, the results were calculated and presented in the format of different charts.

In paper [8] Mugdha and Pradnya talked about the Hadoop framework which was used as a basic structure to do the process of sentimental analysis. Example, reviews of say

washing machines of some XYZ company extracted from e-commerce websites. They used a tool kit in NLP for the preprocessing of the dataset. In the end, the sentimental clarification was done on the basis of Naive Bayes theorem which indeed reduced the comments to the respective polarities.

In paper[9], the authors suggested a way to classify the sentimental reviews from a phone. They were organized into various categories of service, name, and functional standards. After this classification, they managed a new type of tagging which escaped the sentiment from all the furnished reviews. At the time of value extraction, a polariser distinguished the ranking of various sentiments by giving them a specific value, say 1 for positive sentiment. They were able to showcase an effective dashboard with a lot of charts and smiley in the profile picture of the mobile users. This method carried out sentiments easily.

Paper [10] by Kamal focuses on the fact of compiling and identifying opinions. An extraordinary ML technique has been taken into account while mining the words. NLP was also used to combine and skin the sentimental data off. At the end the polarities were determined for different objective and subjective structures. In the end the results of polarity were showcased in a graphical format. Also, specific color criteria were given to each feature visualization.

While paper [11] which was our major inspiration focused on the sentimental analysis of hotels. It had many factors from sanitation to 5-star ratings. A lot of feedback was taken into count. The accusation was meant to be correct as they used a Latent approach for the visualization. The visualizations were based on discrete polarities and summarized the whole structure. It consisted of plenty of different nodes for different functions that were applied to create a polar result. They tried to add various functions of listing the functions needed by the user. Also, the polarities here have different colors to signify the sentiments.

This paper focused on sentiments related to higher education in the United States. The study was based on reviewing the old structure to the new one. The results focused on the true sentimental opinions of the parents regarding the old structure. The geographical map graph was used to see the size of the circle's sentiments in different states and cities. Resulting in, study into visualizing the current trend of the need of change in the education system and working for a good quality of education.[12]

In paper [15] a comparative aspect-based analysis is seen between different algorithms and approaches. The ML algorithm is used to see the major vow between a negative and a positive approach. It overall concluded to an aspect-based sentimental analysis structure which is a boon for future technologies. Also, a new Bayesian estimation is used in the reference which is more reliable and useful than the ones used

in prediction previously. The stock exchange predictions can be a boon to society if done truly and hence, can help the customers also.[16]

Concluding, many papers gave their respected opinions and reviews regarding different products. Therefore, the given system tries to utilize the given visualization techniques and also offers different visualization techniques that can be worthy and useful in structuring the purpose where one can efficiently analyze the sentimental feedbacks. Moreover, interaction with the structure says, choosing a restaurant and extracting its discrete sentimental ratings. If the rating turned out to be neutral the customer may not enter a restaurant. the structure is built on the choices of users. Hence, the structure built describes plenty of customer’s problems with swift results.

3. PROPOSED STRUCTURE

The unified structure tries to furnish the visualized methodology for foodies and restaurants for extracting the appropriate restaurant details according to their needs. The system comprises 2 important parts combined to solve the problem statement: Sentimental Analysis of comments and Visualization using Tableau visual techniques (see figure:1). The proposed system is built on for three steps: to manage the skinned dataset then making an analysis of sentiment out of that and the last part halts the search with charts of analyzed data in tableau which is the visualization phase of our data. The method comprises visualization on Tableau [13] with Python for analytics.

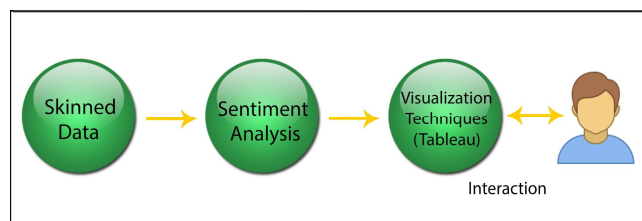


Figure:1: Arranged Architecture

A. Skinned Dataset

While starting with research, the dataset used for the restaurant reviews is from Zomato, which is an open dataset. The food review dataset skinned is in a JSON format that can be easily accepted by the computer. The skinned dataset is extracted from a single large dataset. The computation of the dataset can be described through two descriptive tables (see Table 1) comprising of Review Description while the other table (see Table 2) describing, different cuisines and amenities study revolves around 4 major credits: cuisines, restaurant names, localities, and reviews from the dataset.

Table 1: Review Structure

Features	Description
Resturant Reviewed	Name of The Resturant
Locality	Area where the resturant is situated
Review	Text of Review
Resturant's Address	Address of Resturant
Resturant's Ph. No.	Phone number of Resturant

Table 2: Customer Description Structure

Features	Description
Cuisine	Name of The Cuisine
Star Rating	Rating of Food (out of 5)
Amenities	Like wifi, Parking, etc.

B. Analysis of Sentiments

Analysis of different sentiments is also termed as opinion mining which signifies the gripping of text analysis and processing the language. Classifying the discrete opinions on discrete subjects which binds to service or product. We did an analysis of sentimental opinions of reviews of various restaurants on Zomato Pune to direct the restaurant rating as neutral(0), positive(+1), or negative(-1). Python is a high-level language with a dynamic semantics used by many investigators and scientists for analyzing different human reviews. Therefore, we used IDLE Python 3.7 to arrange the summary of more than 50 restaurants into positive, negative, or neutral. The two major packages found in Python to perform the sentiment purpose are SpaCy and TextBlob, which we used in our analysis. The sentiment package requires a pre-installation of a 64-bit Python stack in the system. After the installations, Python estimated the criteria which are used in the Tableau table of the function of polarity.

C. Tableau Visualization

“Forming a sentimental model is termed as Visualizing” was coined by Spence. [14] The Number of visualizing techniques is bound to provide a visual skinned dataset structure supporting users to summarize plenty of attributes related to different restaurants and the food being served there. Packed bubbles, Bar chart, and Discrete word-cloud summarize the analysis solving the problem statement.

- Packed Bubbles are representations of localities with the most top-rated restaurants.
- Bar Chart shows the percentage of positive reviews in a specific locality to find the high rated restaurants.
- Word-cloud used to visualize the most frequent words used to describe a restaurant. It also signifies the number of restaurants getting more positive ratings.

The three charts are designed to signify the statement furnishing the given methods:

- Meta overview using three visualization techniques of restaurant reviews.
- Interactional service to choose a restaurant and find whether it vows a negative or a positive customer aspect.
- Detailed information like Breakfast, Card- Cash Facility, Timings, etc. detailed information after choosing a restaurant in a specific locality.

4. RESULT

The following subsections will describe the visualization techniques being used to visualize the skinned dataset.

A. Determine the sentiment percentage of different restaurants.

In order to get a general look at the most positively reviewed restaurant, Word cloud is used with different colors shown in figure.2. The text shows different restaurant names have been assigned different sizes according to the positive review feedback percentage. Different colors show different levels of percentage (100 to 0) and so the different sizes. As demonstrated the Word with the biggest cloud is having the highest percentage of positive reviews while the restaurant with the smallest one has the least positive bent. As shown in the figure: Pokket Café has the highest number of positive reviews.

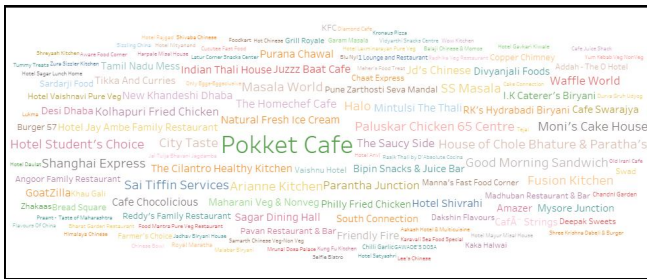


Figure 2: Word Cloud of Restaurant Names

B. Identify the top-rated restaurant in different localities.

The dataset includes more than 30 localities with more than 50 restaurants from Zomato Pune, from a total of 12,000 records. The packed bubble shows the bigger size for more number of highly positive reviewed restaurants. The same chart can also be considered as an area with the most number of amazing cuisines served (see figure.3). As shown in figure Kothrud and Wakad have the most top-rated restaurants. The user can click on the bubble to get the visualized restaurant list for the concerned area.

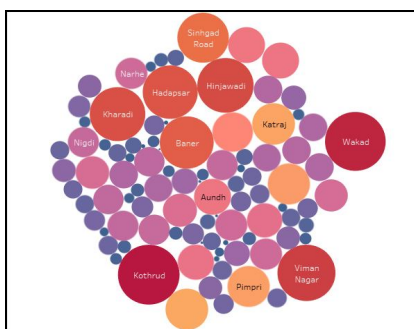


Figure.3: Packed Bubbles of Localities

C. Finding a restaurant review rating within a specific locality

We selected a single locality out of 191 localities from the record and tried to publish the results of all the restaurants present in that locality. Here, (in figure.4) you can see the locality selected is Koregaon Park has more than 15 restaurants and the bar shows the percentage count of positive reviews for the same.

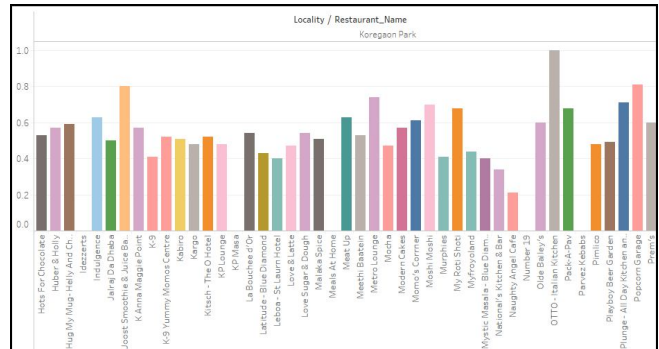


Figure.4: Bar chart of Sentiment Ratings

5. CONCLUSION

In the new digital era where large amounts of datasets are budding on the globe, to analyze the same and extract the result is getting popular. While building the paper, we studied a lot of research in a statistical manner to understand the need and greed of visualizing the dataset. For the same, a large number of researches have been taken into count. Using Python integrated with Tableau, we tried to represent about 1000 reviews from Zomato's restaurant using visualization. Various techniques to visualize the dataset have been used for saving valuable user's and restaurant owner's time in the term of knowing the major quality of reviews of the restaurant is whether positive, neutral, or negative. Users can easily compare the highest-rated restaurant in his/her locality. Also, he/she can find localities with the most suitable cuisines restaurants.

Moreover, users can easily compare different restaurants. Some of the results obtained may not be up to the mark and can make the search ineffective and thus, taken as neutral(0). The logic behind the same is the lexicon approach being followed in Python. Hence, major abrupt or invalid categorization was due to the dialect or argot words that are used in the feedback reviews which we analyzed since these words are not covered in extensively usual lexica.

6. FUTURE SCOPE

Appendixes, The system created by the authors will be clubbed with a discrete restaurant system reviewed by different bloggers. Also, conducting a study of the system to extract the proper outcomes. Furthermore, authors are willing

to expand the design into an adjustable system structure of simply extracting and ranking of the best-given reviews. Also, working on other aspects to provide positive service and remove the jargon words for a proper study. This can be done by adding new words to the lexicon. Including more semantic features in reviews and mining can upgrade the results.

REFERENCES

1. Dave, Kushal, Steve Lawrence, and David M. Pennock. "Mining the peanut gallery: Opinion extraction and semantic classification of product reviews." In *Proceedings of the 12th international conference on World Wide Web*, pp. 519-528. 2003. <https://doi.org/10.1145/775152.775226>
2. Liu, Shixia, Weiwei Cui, Yingcai Wu, and Mengchen Liu. "A survey on information visualization: recent advances and challenges." *The Visual Computer* 30, no. 12 (2014): 1373-1393.
3. 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. <https://doi.org/10.1109/CAIS.2019.8769589>
4. Godwin, Allison. "Visualizing systematic literature reviews to identify new areas of research." In *2016 IEEE Frontiers in Education Conference (FIE)*, pp. 1-8. IEEE, 2016.
5. Salameh, Hani Bani, Ayat Ahmad, and Ashraf Aljammal. "Software evolution visualization techniques and methods-a systematic review." In *2016 7th International Conference on Computer Science and Information Technology (CSIT)*, pp. 1-6. IEEE, 2016.
6. Mikhailov, Sergey, Mikhail Petrov, and Birger Lantow. "Ontology Visualization: A Systematic Literature Analysis." In *BIR Workshops*. 2016.
7. Gundla, Apurva V., and Manisha S. Otari. "A review on sentiment analysis and visualization of customer reviews." *International Journal Of Engineering And Computer Science* 4, no. 10 (2015). <https://doi.org/10.18535/ijecs/v4i10.11>
8. Jinturkar, Mugdha, and Pradnya Gotmare. "Sentiment analysis of customer review data using big data: a survey." *International Journal of Computer Applications* 975 (2016): 8887.
9. Bhatt, Ashutosh Kumar, Durgesh Pant, and Richa Singh. "An analysis of the performance of Artificial Neural Network technique for apple classification." *AI & society* 29, no. 1 (2014): 103-111.
10. Kamal, Ahmad. "Review mining for feature based opinion summarization and visualization." *arXiv preprint arXiv:1504.03068* (2015).
11. Chen, Yu-Sheng, Lieu-Hen Chen, and Yasufumi Takama. "Proposal of lda-based sentiment visualization of hotel reviews." In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pp. 687-693. IEEE, 2015.
12. Qiu, Robin G., Ramya R. Ravi, and Lawrence L. Qiu. "Aggregating and visualizing public opinions and sentiment trends on the US higher education." In *Proceedings of the 17th International Conference on Information Integration and Web-based Applications & Services*, pp. 1-5. 2015. <https://doi.org/10.1145/2837185.2837261>
13. M. Tory and V. Setlur, "Do What I Mean, Not What I Say! Design Considerations for Supporting Intent and Context in Analytical Conversation," 2019 IEEE Conference on Visual Analytics Science and Technology (VAST), Vancouver, BC, Canada, 2019, pp. 93-103, doi: 10.1109/VAST47406.2019.8986918.
14. Spence, Robert. *Information visualization*. Vol. 1. New York: Addison-Wesley, 2001.
15. Maganti Syamala et al. A Deep Analysis on Aspect based Sentiment Text Classification Approaches. *International Journal of Advanced Trends in Computer Science and Engineering* vol.8 no.5, pp.1795-1081, September- October 2019. <https://doi.org/10.30534/ijatcse/2019/01852019>
16. Lakshmana Phaneendra Maguluri et al. A New sentiment score based improved Bayesian networks for real-time intraday stock trend classification. *International Journal of Advanced Trends in Computer Science and Engineering* vol.8 no.4, pp.1045-1055, July-Aug 2019. <https://doi.org/10.30534/ijatcse/2019/10842019>