**Visualization of Review through Social Media**

**Contents:**

1. Problem Statement…………………………………………………………..2
2. About Company……………………………………………………………..3
3. Emergent Technology……………………………………………………….4
4. Approach……………………………………………………………….........5
5. Implementation……………………………………………………………...6
6. Visualizations………………………………………………………………..8
7. Division of Work…………………………………………………………...11
8. Conclusion………………………………………………………………….12
9. Future Development………………………………………………………..13
10. References………………………………………………………………….14

**Abstract**

We propose a proof – of – concept to perform sentiment analysis on user’s post or messages posted via social networking websites, such as Facebook. The POC for ‘Cakemporos’, a Mumbai based startup primarily focused on delivering bakery and confectionary products online. Cakemporos maintains an active Facebook page with a response time of less than five minutes. Many customers post on the wall of Cakemporos’s Facebook page or send private messages to provide suggestions or file complaints. This inherent form of feedback can be used by the company to analyze their services.

1. **Problem Statement:**

* Creating a visual representation by performing sentiment analysis of a Facebook page’s chat conversations with its subscribers.
* Visualization would include the locations of the users along with their satisfaction score which can be determined by using a Natural language processing algorithm.

1. **About Company:**

The company is a Mumbai based startup named Cakemporos founded in 2014 by two college students from St. Francis Institute of Technology. It is a Mumbai based online vendor of bakery products. It lists the best bakeries in town based on reviews, experiences and recommendations. Cakemporos maintains an active Facebook page with a response time of less than five minutes. Many customers post on the wall of Cakemporos’s Facebook page or send private messages to provide suggestions or file complaints.

1. **Emergent Technology:**

Tableau is a complete Data Analytics platform enabling anyone to gain insights from raw data. Tableau helps in visualizing data and has capability to easily connect to wide array of data sources.

**Functionality of Tableau:**

* Tableau can format data in a pictorial or graphical view to easily spot patterns or correlations between various data points. It can connect to third-party tools to pull data and is user-friendly with features like simple drag and drop or drop-down menu.
* Tableau automatically updates information in real-time into its system, easily refreshing data.
* Tableau seamlessly integrates with existing security protocols maintaining security of client database with pass-through data connection permissions and row-level filtering.

**Business Value:**

* Many organizations rely on up-to-date data to make critical decisions. Tableau’s automatic data handling in real-time is what makes it lucrative for both large and small businesses.
* Tableau server offers collaboration and security for data visualizations. The data can be retrieved from any source and shared within the organization in a secure way using Tableau server.

1. **Approach:**

Traditional feedbacks methods like survey form, google form are inefficient and demand significant time from customers. Only 9% of response rate is observed on an average in past decade for such feedback forms. The traditional one way feedback systems are generic and are not connected to data or individual customer experience. Our concept of implicit feedback through social media platform paves the way for genuine and contextually appropriate feedback from the users.

For instance, if a customer is dissatisfied with the service provided by the company, he/she can post on the page. This helps us to collect the positive and negative feedbacks and target the areas where improvement is required. Since users can post about their experience immediately, they are considerably genuine and hence we can conclude that the quality of feedbacks obtained from this method is better than the traditional ones. By using this analysis, the company can also provide promotional offers to customer in order to make up for their poor service.

Measure of success of this concept relies on number of genuine feedbacks received from the users. Response rate are increased by taking implicit feedbacks. Other important criteria for measuring success of our concept is to detect the false positives generated by the Natural language processing and successfully handle it. Also, to distinguish between false and true feedbacks, we will validate the user’s Facebook URL which will serve as an unique identifier. These all functionality improve the results significantly.

Customer Service has been proved to be the most integral section of business. Feedback is the prime most way to understand customer satisfaction. Our concept let's company understand the customer`s opinion towards service and aids to find out the sectors that need improvement. This whole process plays major role in creating customer value. It also provides the context of feedback which can be further used to provide promotional offers to customers.

Implementation of this concept requires investment in different components of the software system. A commercial tableau software, whose main purpose is to visualize the data and help in pattern finding is to be purchased. A server would be required to handle the traffic from Facebook. This can be deployed on the current server which hosts the website. By visualizing the location of the customer and their ‘likert’ scale value on a map, we can understand areas where customers require more attention or better quality of service.

Finally, we need to integrate all the individual components to start gathering the data for further visualization. The Graph API delivers the messages to the server through a webhooks service which is available on the Facebook app’s dashboard. The server receives the data and invokes the NLP algorithm which parses the data into sentences and performs data cleaning operations.The final dataset consists of customer’s name, sentiment score and location information. This dataset, in csv form, is then imported into Tableau. Tableau automatically turns the location information into rich, informative map plots with 16x interactive zoom. The sentiment scores will be color-coded and plotted onto the map. Different types of visualization on this dataset will be explored which would help the startup to better understand its services. This simple, yet highly effective technique would be a paradigm shifting approach for any feedback system.

1. **Implementation:**

The basic algorithm used for performing Natural Language processing is Support Vector Machines:

1. Read the csv file using pandas.io library and store it as a Dataframe.

2. Classify the attributes as samples and features to apply it to Support Vector Machines. Features are also referred as predictors and samples as responses.

3. Create an object of Classifier imported from scikit learn library.

4. Using fit function, fit the dataframe to tree model. Fit function finds patterns in the given dataset.

5. Predict the behavior of model using predict function with test data.

6. After cross validation use the test dataset and predict again.

7. Calculate accuracy using score function.

For Text conversion, we first removed all the stopwords using the nltk toolkit. Later on, we converted the words to vectors and found the occurrences of each meaningful words using CountVectorizer. This technique is referred to as ‘Bag of Words’ technique which does not consider the order of occurrences of words or their meaning.

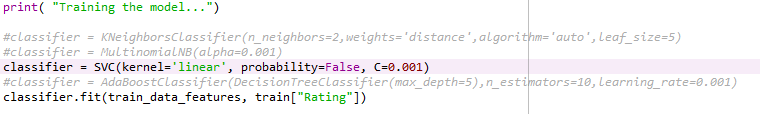
We tried implementing model using other algorithms such as Decision trees, K Nearest Neighbors, Adaboost with Decision Trees but SVM gave better results and accuracy than rest others. Given a set of training examples, each marked as belonging to one or the other of two categories, an SVM training algorithm builds a model that assigns new examples to one category or the other, making it a non-[probabilistic](https://en.wikipedia.org/wiki/Probabilistic_classification) [binary](https://en.wikipedia.org/wiki/Binary_classifier) [linear classifier](https://en.wikipedia.org/wiki/Linear_classifier). An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible.

The advantages of support vector machines are:

* Effective in high dimensional spaces.
* Still effective in cases where number of dimensions is greater than the number of samples.
* Uses a subset of training points in the decision function (called support vectors), so it is also memory efficient.
* Versatile: different [Kernel functions](http://scikit-learn.org/stable/modules/svm.html#svm-kernels) can be specified for the decision function. Common kernels are provided, but it is also possible to specify custom kernels.

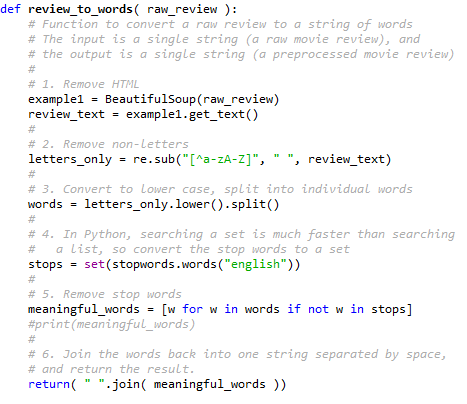
The disadvantages of support vector machines include:

* If the number of features is much greater than the number of samples, the method is likely to give poor performances.
* SVMs do not directly provide probability estimates, these are calculated using an expensive five-fold cross-validation

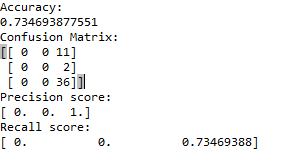


The above code snippet shows how the model is trained using SVM.

The code snippet below shows the text conversion technique used.



**Output:**

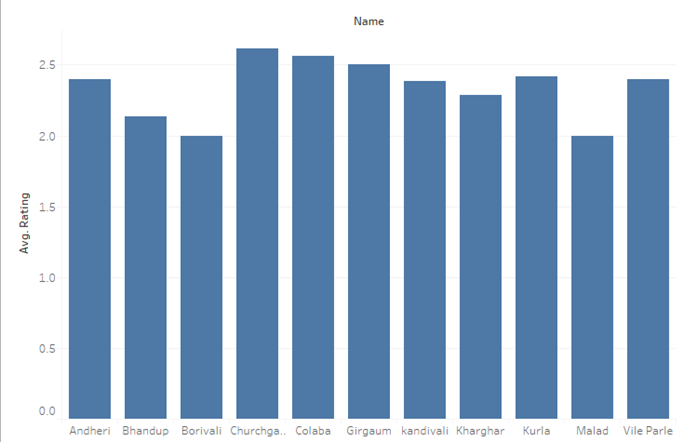


1. **Visualizations:**

By using tableau we have come up with the below visualizations for our dataset.

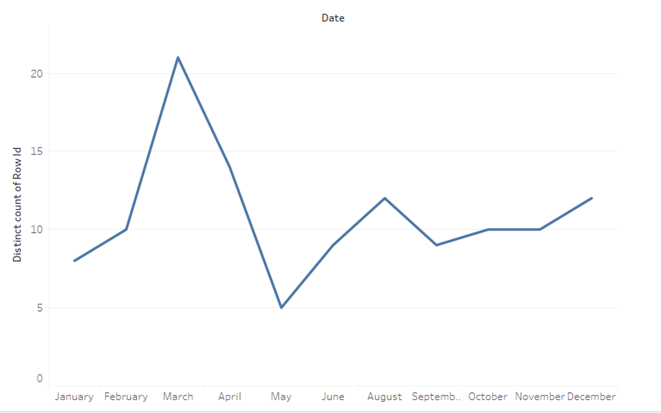
1. Comparison of ratings given by the customer and location.

* This graph shows the ratings given by the customers from a particular place.
* We can observe that in particular places, the ratings have reduced.
* This can be possible due to low customer service in that area and hence we can target these places in order to improve on the service and have better customer satisfaction.



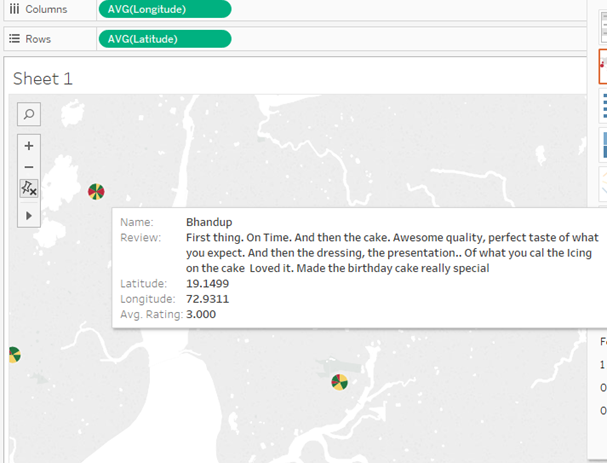
     2.  Comparison of number of customers per month.

* This is a line graph which shows the number of customers per month.
* Due to low customer service, it is possible that the number of customers reduce in a particular time of the year.
* From the below graph we can see that in the months of January, May and December the number of customers have reduced.



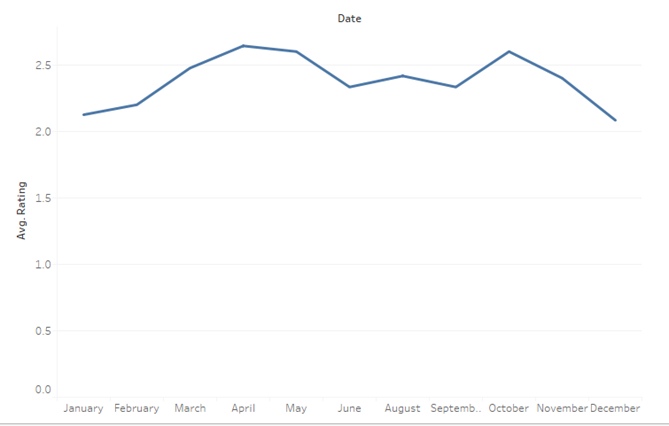
    3.    Symbol Map of the entire dataset.

* By using tableau we can plot locations of the service using their latitude and longitude and also observe the the ratings and reviews given by particular customers to the company’s quality of service.
* It gives us an overall idea of the areas the company need to target in order to increase their business.



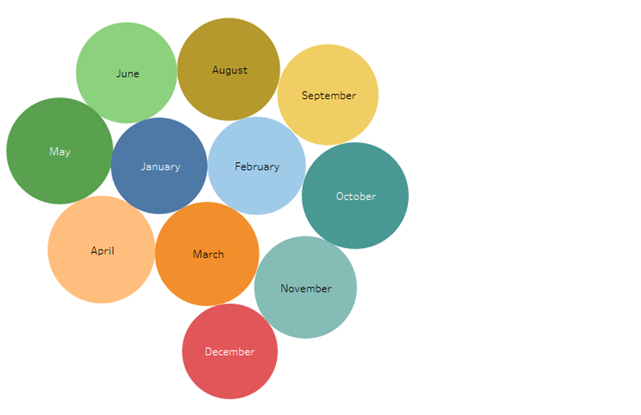
4.   Comparison of ratings in a particular month.

* This is a line graph which shows ratings given by customer in a particular month.
* Sometimes due to the occurrence of festivals in a particular month, the business increases but can also lead to reduced quality of customer service



   5. Comparison of ratings in a particular month.

* This is a bubble graph of the above case. The larger the size of the bubble, the greater is the ratings given by the customer.



1. **Conclusion:**

Different types of visualization on this dataset were explored which would help the startup to better understand its services. This simple, yet highly effective technique would be a paradigm shifting approach for any feedback system.

The visualizations obtained based on map locations suggest that the customers from particular area aren’t satisfied which will in turn help company to dig down and find the actual cause of dissatisfaction. Company can extend offers to attract customers based on occasions and trend observed in the purchases made.

This idea of analyzing customer reviews and ratings can help in betterment of the business of the company.

1. **Future Development:**

We can modify the NLP algorithm used that is ‘Bag of Words’ to ‘Continous Bag of Words’ which preserves the context and grammar of text. This approach will give better results for reviews as the reviews are one line or just a group of words.

Extracting the data from Social media like Twitter and performing the same analysis for better results.We can create customized dashboards in Tableau to display data.

This proof of concept can be extended to different organizations who need customer feedback analysis. By modifying the dataset as per the customer requirements we can obtain different patterns, visualizations which will help the company with business.

1. **References:**

[1]Tableau customer stories,

<https://www.tableau.com/stories/customer/optumhealth-understands-its-customers>

[2] Timextender,<http://www.timextender.com/wp-content/uploads/2015/12/Tableau.pdf>

[3] Bardess business value,

<http://www.bardess.com/solutions/data-discovery/tableau/business-value>

[4] Scikit-learn SVM, <http://scikit-learn.org/stable/modules/svm.html>

[5] CountVectorizer,

<http://scikitlearn.org/stable/modules/generated/sklearn.feature_extraction.text.CountVectorizer.html>