

Sentiment Analysis of Customer Feedback on Restaurants

Spoorthi C. B.E., (M.Tech)¹
CS & E dept, AIT College
Chikkamagaluru, India

Dr. Pushpa Ravi Kumar B.E., M.Tech. Ph.D²
CS & E dept, AIT College
Chikkamagaluru, India

Mr. Adarsh M.J. B.E., M.Tech³
CS & E dept, AIT College
Chikkamagaluru, India

Abstract: Sentiment analysis is a huge volume increasing at a humongous rate everyday which has made it almost impossible to evaluate the data manually. In Social media, twitter, restaurant site people share their opinion as in a huge number of their prevalence. In order to make the process of analyzing the text automatic there are various machine learning techniques that could be applied. The data set is for those enthusiasts who are willing to play with text data and perform sentiment analysis or text classification. The huge quantity of data in textual is generated every day has no value unless processed. The text data issue can be resolved by adopting Data mining technique using r tool. Our experimental work intends to adopt Naïve Bayes classifier is data mining techniques for the effective prediction of text data. This data set consists of actual reviews from real people. So this data set will give a real time experience as to how to deal with textual data.

Keywords: Sentiment analysis, Social media, Naïve Bayes classifier, Restaurant reviews.

I INTRODUCTION

Recently there has been number of restaurants when you are on the lookout for a new place to eat, what is the best way to find a great restaurant. Ask someone who's been there, of course. If you don't have someone to personally ask, then you can always turn to online reviews. Customers take many factors into consideration that when deciding where to eat. It's not just about how great the food tastes but how good the service is, how polite the employees are, and how well maintained the facilities are. The truth is, consumers are trusting advertising less and less and turning to reviews to find out what dining at a restaurant is really like. Customer having testimonials will give the potential and also the customers assurance that they may have a great experience.

Customers want to know what to expect when trying a new restaurant. And who better to tell them than a previous customer. The more individuals hear about your restaurant, the more inclined they will be to dine there. Now it is known that people are now incline to turn to customer reviews first than to decide where to eat. Don't let a lack of reviews for your restaurants prevent you from standing out. Collecting the recommendations by making the customers easy to talk about how great their experience was in choosing.

Restaurant Review

It is simple, people believe each other. Customer does not believe directly when choosing a restaurant or hotel, they believe when their buying a phone, car or

clothes from an online. They believe that their reviews are pragmatic and that they can know what to expect while reading them. Although a negative review can come as a shock for owners, they must know that even the best get bad reviews and that the whole sum is the real picture of what they offer. So, restaurant, bar or accommodation owners need to encourage people to make reviews and share their experience and doing so they practically are saying –we do quality stuff and our service is always on high level. Your opinion matters to us! . Online reviews make it possible for people to say their opinion from their home, on the back seat of a car while driving home without having to confront with anybody.

The most important is the sum of reviews that makes a list on which one can assume how much a restaurant for example is popular. The review is compounded of grades for service, ambient and cleanliness. The influence can be huge. It has shown that a rise of grade for one can increase revenue from 5 to 9 percent what can have a positive impact on the whole firm. This kind of visibility of restaurants, bars and accommodation has given the possibility for those smaller and on less attractive locations to reach large number of guests. Today it is not important where you are or the history of your place, it is important what is the level of your service.

A. Social Media

Make suring that we all using some sort of social media and having a page on facebook makes our venue rateable and courages from a people to tag more people when they all having their rating food. People will post on some sorts

B. Google

Now a days it has become a number one position and second position is a food online reviews. Food sites focus more on reviews.

C. Yelp

Yelp ranks has a second 45.18 percent followed by some many people and by trip advisor. The popularity is getting more on third party review sites like google, facebook, yelp, and trip advisor is driven by customers' genuine desire to engage with their businesses.

II LITERATURE SURVEY

J. P. Schomberg, O. L. Haimson, G. R. Hayes, and H. Anton-Culver [2] has proposed the "Supplementing public health inspection via social media" Mining publicly-available crowd sourced data to develop a surveillance

method for tracking foodborne illness risk factors gives health inspectors an improved ability to identify restaurants with greater odd so flow health code ratings and violations outside of the normal inspection window.

A. Sadilek, S. Brennan, H. Kautz, and V. Silenzio Nemesis [3] has proposed “Which restaurants should people avoid today?” Computational approaches to health monitoring and epidemiology continue to evolve rapidly. The proposed work presented an end-to-end system, nEmesis, that automatically identifies restaurants posing public health risks. Leveraging a language model of Twitter users’ online communication, it makes the people to nEmesis finds individuals who are likely suffering from a foodborne illness from a colony. People’s visits to restaurants are modelled by matching GPS data embedded in the messages with restaurant addresses.

C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky [5] has proposed “Stanford corenlp natural language processing toolkit” the proposed work describe the design and use of the Stanford Core NLP toolkit, an extensible pipeline that provides core natural language analysis. This tool kit is quite widely used, both in the research NLP community and also among commercial and government users of open source NLP technology. The method suggest that it follows from a simple, approachable and easy design, which can be straightforward interfaces, the inclusion of robust and in the good quality analysis components, and not to requiring use of a large amount of associated baggage. The approached method defines the design and development of stanford Core NLP, and gives the common core natural language processing steps, from the tokenization.

K. Lee, A. Agrawal, and A. N. Choudhary [6] has proposed “Mining social media streams to improve public health allergy surveillance” Allergies are one of the most common chronic diseases worldwide. One in five Americans suffer from either allergy or asthma symptoms. With the prevalence of social media, people sharing experiences and opinions on personal health symptoms and concerns on social media are increasing. Allergy is the fifth most common chronic diseases in the United States¹. The complexity and severity of allergic diseases are increasing worldwide. One in five Americans have either allergy or asthma symptoms. In 2012, 7.5% of adults (17.6 million adults) and 9% of children (6.6 million children) were diagnosed with hay fever. Continuous use of allergy medication can worsen patients’ health conditions and lead to side effects and other serious medical complications.

K. Lee, A. Agrawal, and A. Choudhary [7] has proposed the “Real-time disease surveillance using twitter data: demonstration on flu and cancer” Social media is producing massive amounts of data on an unprecedented scale. Here people share their experiences and opinions on various topics, including personal health issues, symptoms, treatments, side-effects, and so on. The proposed work do the publicly available social media data an invaluable resource for mining available and wanted actionable healthcare in the media. In this paper, we describe a novel real-time flu and cancer surveillance system that uses spatial, temporal and text mining on Twitter data. The

Internet is usually the first place people turn for health information. People search for a specific disease, symptoms, and appropriate medical treatments, and often make decisions whether they should go see a doctor based on the search results.

III METHODOLOGY

This proposed work is to predict the text automatically based on the data set values stored by using the r tool. By using the training data set values it is possible to predict the text data using our classifier called naïve bayes using algorithm.

The Fig 1 depicts the architecture of the proposed model used in the prediction of sentiment analysis. It consists of 3 steps

A. Data Collection

In this step data is taken out from kaggle in a recognized format. Missing fields are evacuated in this process & thus the data is transformed. Sentiment Analysis can be considered a classification process. There are three main classification levels in sentiment analysis document-level, sentence-level, and aspect-level sentiment analysis. Level of document it aims to classify an opinion document which as a positive or negative opinion expression. It considers the full document as a basic information unit.

B. Data Preprocessing

The collected raw data of restaurant reviews consist of large number of attributes and also there will be missing values. The reducing the attributes is required, extracting the required attributes is also much essential. So in order to get importance of the each variable or attributes migrtr algorithm is applied. Migrtr algorithm which selects the attributes based on predictor, here predictor considered restaurant review. Feature or Attribute extraction is done using migrtr algorithm. In detail steps working of migrtr algorithm.

In Data cleaning once attributes are removed, filling the missing values, removing inconsistent data measuring the central tendency for the attribute such as mean median, quartile is done. In data preprocess the data is cleaned and the extracted data before analysis. Non-textual contents and contents that are irrelevant for the analysis are identified and eliminated.

C. Sentiment Analysis

The reviews sources are mainly review sites. Sentiment analysis is not only applied on product reviews but can also be applied on stock market, news articles, or political debates. In political debates for example, we could figure out people’s opinions on a certain election candidates or political parties. The election results can also be predicted from political posts. The sites like social media and micro blogging sites are taken a very good source of information because many people share and discuss their opinions about positive and negative opinion freely.

D. Classification

The lexicon-based approach is to finding the opinion mining which is used to analyze or to predict the text. There are two methods in this approach. The dictionary-based approach which depends on finding opinion seed words, and then searches the dictionary of their synonyms and antonyms. The corpus-based approach begins with a seed list of opinion words, and then finds other opinion words in a large corpus to help in finding opinion words with context specific orientations. This could be done by using statistical or semantic methods.

Data mining has got two most frequent modelling goals – classification & prediction. Classification model classifies discrete, unordered values or data. In this prediction process, the classification techniques utilized are, naive bayes classifier.

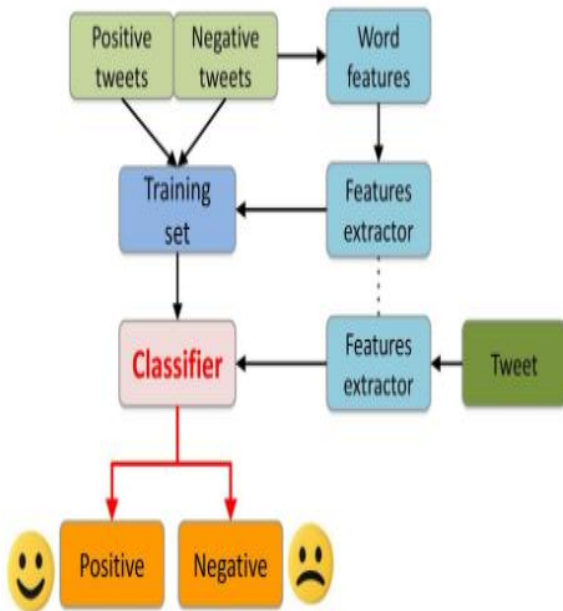


Fig1: Architecture of the proposed model

A. Naïve Bayes

It is one of the popular classification techniques of algorithms used in data mining. It is a probability classifier. It links the attributes mutually & is dependent on the number of parameters. The principle here is that the variables provided are independent. It generates accurate results with appropriate calculation & provides fast results. It is based on Bayes theorem & the formula is,

$$P(\text{label}|\text{features}) = \frac{P(\text{label}) * P(\text{features}|\text{label})}{P(\text{features}) \dots} \dots(\text{eq1})$$

Fig 2 shows the graphical representation of the result.

The graph is generated by using the hchart chart. X-axis is the number data set values and Y- axis represents the Probability of positive or negative with the percentage values.



Fig 2: Calculation of positive, negative and average probability using Naïve bayes

IV RESULT & DISCUSSION

Data analysis is the most crucial part of any proposed work. Data analysis summarizes collected data. It involves the interpretation of data gathered through the use of analytical and logical reasoning to determine patterns, relationships or trends. To examine critically and to bring out the essential elements or give the essence to analyze a data. To examine carefully and in detail so as to key factors, possible results. Following snapshots shows the results obtained in each step of the process.

Fig 3 depicts naïve bayes algorithms with the accuracy vales like positive is 3456, negative is 485, and average is 27 . Fig 4 shows the test cases of the algorithms with different accuracy values like 112. and Fig 5 shows one more test case result like 44 and 28.

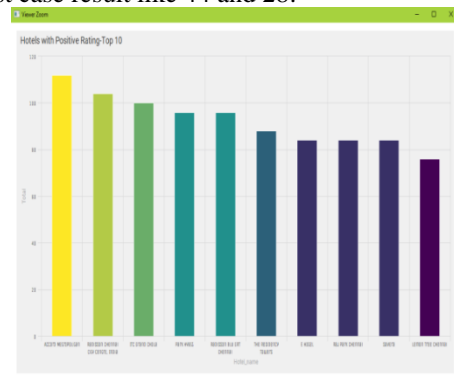


Fig 3: Comparison of reviews with positive and naïve bayes algorithm

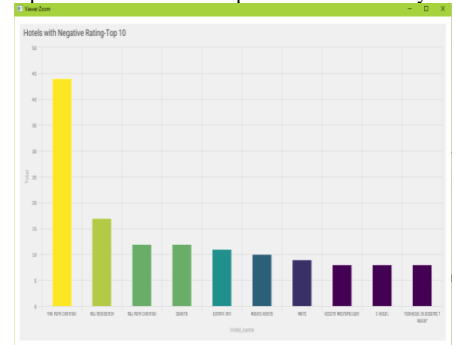


Fig 4: Comparison of reviews with negative and naïve bayes algorithm

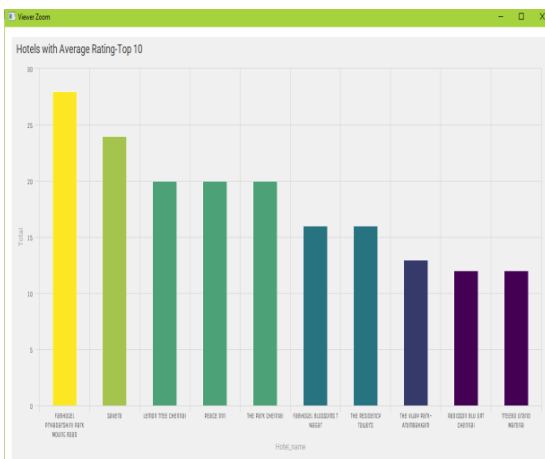


Fig 5: Comparison of reviews with average and naïve bayes algorithm

V CONCLUSION

The proposed work starting from the analysis of different studies provided in the literature, provides a classification of sentiment classification approaches with respect to features/techniques and advantages /limitations, tools for sentiment analysis with respect to the different techniques used for sentiment analysis.

The sentiment classification approaches can be classified in machine learning, lexicon based and hybrid approach. The machine learning approach is used for predicting the sentiments based on trained and test data sets. In our lexicon based approach does not need any prior training in order to mine the data.

REFERENCE

- [1] Mikel Joaristi, Edoardo Serra, Francesca Spezzano "Evaluating the Impact of Social Media in Detecting the Restaurants Violating the Health Norms" In 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM).
- [2] J. P. Schomberg, O. L. Haimson, G. R. Hayes, and H. Anton-Culver. Supplementing public health inspection via social media. *PLoS ONE*, 11(3), 03 2016.
- [3] A. Sadilek, S. Brennan, H. Kautz, and V. Silenzio. nemesis: Which restaurants should you avoid today? In *HCOMP*, 2013.
- [4] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, et al. Scikit-learn: Machine learning in python. *The Journal of Machine Learning Research*, 12:2825–2830, 2011.
- [5] C. D. Manning, M. Surdeanu, J. Bauer, J. R. Finkel, S. Bethard, and D. McClosky. The stanford corenlp natural language processing toolkit. In *ACL (System Demonstrations)*, pages 55–60, 2014.
- [6] K. Lee, A. Agrawal, and A. N. Choudhary. Mining social media streams to improve public health allergy surveillance. In *ASONAM*, pages 815– 822, 2015.
- [7] K. Lee, A. Agrawal, and A. Choudhary. Real-time disease surveillance using twitter data: demonstration on flu and cancer. In *KDD*, pages 1474–1477, 2013.
- [8] A. Lamb, M. J. Paul, and M. Dredze. Separating fact from fear: Tracking flu infections on twitter. In *HLT-NAACL*, pages 789–795, 2013.
- [9] J. S. Kang, P. Kuznetsova, M. Luca, and Y. Choi. Where not to eat? improving public policy by predicting hygiene inspections using online reviews. In *EMNLP*, pages 1443–1448, 2013.
- [10] M. Dredze, M. J. Paul, S. Bergsma, and H. Tran. Carmen: A twitter geolocation system with applications to public health. In *AAAI/HIAL*, pages 20–24, 2013.
- [11] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer. Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, pages 321–357, 2002.
- [12] E. Aramaki, S. Maskawa, and M. Morita. Twitter catches the flu: detecting influenza epidemics using twitter. In *EMNLP*, pages 1568– 1576, 2011.