# SENTIMENT ANALYSIS ON RESTAURANT REVIEWS

## ABSTRACT

Sentiment analysis of customer reviews has a crucial impact on a business's development strategy. Evolution of the internet in the past decade resulted in generation of voluminous data in all sectors. Due to these advents, the people have new ways of expressing their opinions about anything in the form of Google Reviews, Tweets, Blog Posts etc. Sentiment analyisis deals with the process of computationally identifying and categorizing opinions expressed in a piece of text, especially in order to determine whether the writer's attitude toward a particular topic is positive, negative or neutral. Knowing the opinion of customers is very important for any business.

Hence, in this paper, we analyze the reviews given by the customers of the restaurant with help of machine learning classification algorithms. The Modern area of sentiment mining also called opinion mining .Researcher in the area of natural language processing(NLP), data mining, machine learning, Support Vector Machine(SVM) and test the method of sentiment analysis process. This problem can be addressed by an automated system called sentiment analysis and opinion mining that can analyze and extract the users view in the reviews.

## I.   INTRODUCTION

Sentiment Analysis is the way toward deciding if a portion of writing is negative, positive or neutral or on the other hand nonpartisan. This piece of writing could be a tweet, review about a book, film, movie, restaurant and so on. The sentiment analysis is also known as opinion mining, in which the opinions, appraisals, emotions or attitude towards a topic, person or entity are analyzed. The expressions can be classified as positive, negative or neutral For example. "I really liked the garlic noodles of your restaurant "- this is a positive expression. The overall sentiment polarity shows a preference on service in the reviews, which might hint the customers to "self -select" the food they like. Natural

language processing in artificial intelligence applications makes it easy to gather product reviews from a website and understand what consumers are actually saying as well as their sentiment in reference to a specific product. Companies with a large volume of reviews can actually understand them and use the data collected to recommend new products or services based on customer preferences. It tends to be utilized to recognize the client or consumer's mentality towards a brand's crucial factors, for example, tone, context, emotion and so forth. These sorts of reviews are equally important for both the consumers and the betterment of the service. From consumers perspective having a view over that service from other consumers is useful for him to get an overall idea of the product. On the other hand, owners or service providers use sentiment analysis to have view about the acceptance of their products or to analyze customer satisfaction and suggestions. But as we can assume it is a very lengthy and time-consuming approach to go through that huge number of reviews and manually analyze the sentiment of those contents. Using sentiment analysis, the overall result about the opinions and views can be obtained within seconds. It not only gives the owners an idea about the consumers, but it also gives them a better picture of how they stack up against their competitors' company. Restaurant reviews are still in the form of text, customer reviews are included in the text mining category, the results of these data will be classified into two values, positive or negative. for preprocessing review data such as remove stopword, remove punctuation done with the help of Python.

## II.   LITERATURE REVIEW

The major step involved in determining the sentiment of a text. In our approach, we have split the preprocessing part into three major steps.

The first step involves removing the punctuation in the sentences. All special characters like exclamatory mark and quotes are removed by designing appropriate regular expression. The resultant data would be containing only alphabetical characters.

The second step involves removing the stop-words from the reviews. Stopwords are the words which are not used to express any emotion or sentiment but used as connectors or articles in the English language. This includes words like and, with, of , the. Natural language processing (NLP) techniques like Lexical analysis, syntactic analysis, semantic analysis,

disclosure integration, and pragmatic analysis are applied on the dataset to identify and remove stop-words.

The semantic analysis step generally removes the stop-words like not as well. But, in opinion mining, the presence/absence of the word not plays an important role. For example, the review says The crust is not good. The removal of stop-words will result this sentence into crust good. Thus, a negative opinion is turned into positive. To avoid this problem, we have modified the semantic analysis step in NLP and made sure that such stopwords are not being removed in the process.

The third step in preprocessing is to convert the original words to their root words. Root words are the words without prefix or suffix. For example, love is the root word for the words loving, loved, loves, etc. As we are interested only in actual opinion/sentiment rather than English grammar, such conversion eases the job. The Porter Stemmer algorithm is applied for converting all words in the dataset into root words.

# III.  METHODOLOGY

**RESEARCH METHODOLOGY**

We developed sentiment analysis system using the standard machine learning approach as explained in the background section.

**OBJECTIVE**

This project focusing on the estimation of the polarity of the sentiment evoked by an  text through input box. To implement an algorithm for automatic classification of text into positive, negative or neutral. Sentiment Analysis to determine the attitude of the mass is positive, negative, neutral towards the subject of interest. It is represented in the form of pie chart.

**PROBLEM STATEMENT**

To provide a Sentiment Analysis system for customers review classification, that may be helpful to analyze the information  where opinions are highly unstructure and are either positive or negative.

**EXISTING SYSTEM**

The content of user generated opinions in the social media such as face book, twitter, review sites, etc are growing in large volume. These opinions can be tapped and used as business intelligence for various uses such as marketing, prediction, etc. Generally sentiment analysis is used for finding out the aptitude of the author considering some topic. But in our social network sites not implemented Sentiment analysis. Some survey depends on the static sent word dataset to find the sentiment analysis. But we require finding a proper solution to find the polarity of the  micro blogs.

**PROPOSED SYSTEM**

We will collect the unstructured data through the text box. With that data covert the data to lower case and data is processed as follow.

**Pre-processing**

Before the feature extractor can use the reviews to build feature vector, the review text goes through preprocessing step where the following steps are taken. These steps convert plain text of the review into process able elements with more information added that can be utilized by feature extractor. For all these steps, thirdparty  tools  were used that were specialized to handle unique nature of review text. PyCharm is an integrated development environment(IDE) used in computer programming. Specifically for the Python language. It is developed by the Czech company JetBrains. It provides code analysis, a graphical debugger, an integrated unit tester.


Step 1: Tokenization

Tokenization is the process of converting text as a string into processable elements called tokens. In the context of a review, these elements can be words, emoticons, url links, hashtags or punctuations "an insanely awsum…." Text was broken into "an", "insanely", "awsum"…. These elements are often separated by separated by a space. On the other hand, hash tags with"#" preceding the tag needs to be retained since a word as a hash tag may have different sentiment value than a word used regularly in the text.


Step 2 : Parts of Speech Tags Parts of Speech (POS) tags are characteristics of a word in a sentence based on grammatical categories of words of language. This information is essential for sentiment analysis as words may have different

sentiment value depending on their POS tag. For example, word like "good" as a noun contains no sentiment whereas "good" as an adjective positive sentiment.

each token extracted in the last step is assigned a POS

Step 3: Dependency Parsing

For our purposes, dependency parsing is extracting the relationship between words in a sentence. This can be useful in identifying relationship between "not" and "good" in phrases like "not really good" where the relationship is not always with the adjacent word.

| CLASS | REVIEWS |
|---|---|
| positive | Had A Great Experience After so long. amazing Elite serving experience |
| negative | Service is very slow, Worst hospitality, Worst experience... |
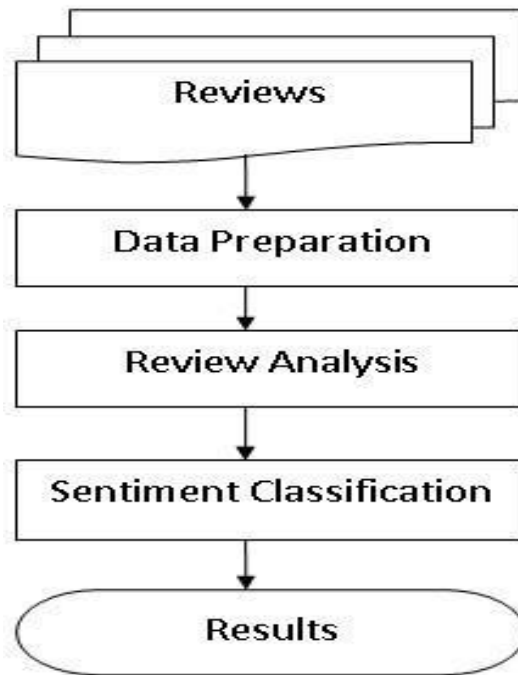
**Fig 1.** Sample Reviews for each sentiment class

# IV. MODELING AND ANALYSIS

**MATPLOTLIB.PYPLOT**

matplotlib. pyplot is a collection of functions that make matplotlib work like MATLAB. Each pyplot function makes some change to a figure: e.g, creates a figure, creates a plotting area in a figure, plots some lines in a plotting are, decorates the plot with labels, etc.

In matplotlib. pyplot various states are preserved across function calls, so that it keeps track of things like the current figure and plotting are, and the plotting function are directed to the current axes (please note that "axes" here and in most places in the documentation refers to the axes part of a figure and not the strict mathematical term for more than one axis)
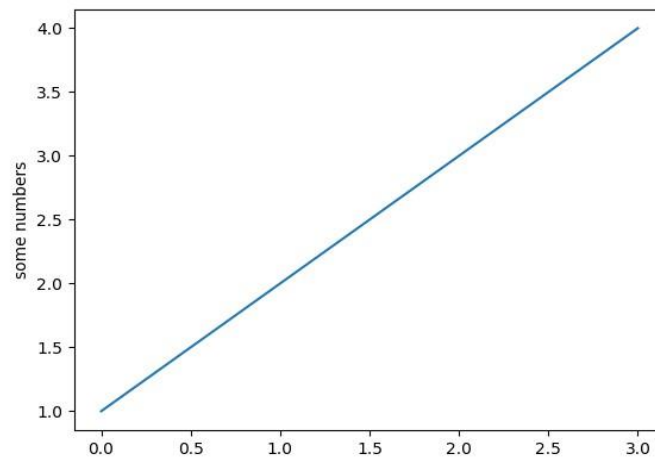
Fig 2. Matplotlib.Pyplot

# V.  RESULTS AND DISCUSSION

## TESTING

```
In [20]: sample_review ='The food is really bad.'

         if predict_sentiment(sample_review):
           print("Positive review")

         else:
           print("Negative review")
```

Negative review

```
In [21]: sample_review ='Food was pretty bad and the service was very slow'

         if predict_sentiment(sample_review):
           print(" Positive review")

         else:
           print(" Negative review")
```

Negative review

```
In [22]: sample_review ='The food  was absolutely wonderful,from preparation to presentation, very pleasing.'

         if predict_sentiment(sample_review):
           print("This is a Positive review")

         else:
           print("This is a Negative review")
```

This is a Positive review

```
In [23]: sample_review ='food was very bad'

         if predict_sentiment(sample_review):
           print("This is a Positive review")

         else:
           print("This is a Negative review")
```

This is a Negative review

7

# VI. CONCLUSION

This document is a study of important techniques to identify sentiment analysis of reviews. In this paper major approaches applicable to identify the sentiment analysis of review text to be analyzed. We cover the maximum out of all existing techniques and also do the comparison of these techniques and use nlp for tokenization, remove stopwords and punctuation marks are also do stemming. And use a supervised machine learning algorithm which is linear regression and naïve multinomial algorithm for classification.