© May 2022 | IJIRT | Volume 8 Issue 12 | ISSN: 2349-6002

SENTIMENT ANALYSIS OF CUSTOMER REVIEWS-USING AI/ML APPROACH

Aaditya Prakash Pillai, Dr. S. P. Chauhan
Department of computer Science and Engineering, School of Computing Science and Engineering, Galgotias University, Greater Noida, India

***Abstract*- Businesses all over the world are nowadays moving to online platforms. This promises the customer the choice to receive the product at their residence (home-delivery), browse through latest offers and new offers from the comfort of their home. The customer not only receives a service but can also give feedback to the company. These are called customer feedbacks. Customers nowadays look up customer feedback and ratings that the service or product receives before they decide on a purchasing decision. Customer reviews can seriously affect the performance of a company. The final verdict about a company's overall quality of service or product is influenced by customer feedback or reviews. Using sentiment analysis, customer feedback can be classified as negative, positive or neutral comments. Sentiment analysis is the computational study of people's opinions, sentiments, emotions and attitudes, it is also known as opinion mining. The aim of the paper is to present the existing approaches applied for sentiment analysis using Machine learning and natural Language Processing technique in a service environment and to investigate the accuracy rates of these algorithms and an app will be developed to help the end user, i.e. the company to receive a very good idea about the customer sentiment about their product and/ or service.**

***Index Terms*- Aspect-based, sentiment analysis, Machine Learning, Natural Language Processing, Streamlit Web application, Support Vector Machine**

I. INTRODUCTION

Sentiment analysis (SA) and opinion mining include studies to analyses people's sentiments, opinions, evaluations, attitudes, and emotions from written language, and as one of the active research areas in natural language processing and is also widely studied in data mining [1].

The customer reviews can seriously affect the performance of a company. The final verdict about a company's overall quality of service or product is influenced by customer feedback or reviews. [1,5] These customer feedbacks are a rich-source data source for companies to analyze has and understand their customers, their requirements and their level of satisfaction or dissatisfaction with the company's services.

In order to make more detailed and productive strategies for product creation and marketing, business participants must be aware of the prevalent consumer expectations and consumer input is available on more online platforms like websites, social media platforms such as Twitter, Instagram and Facebook [3]. Using sentiment analysis, customer feedback can be classified as negative, positive or neutral comments. [2], we find that our model detects business failures.

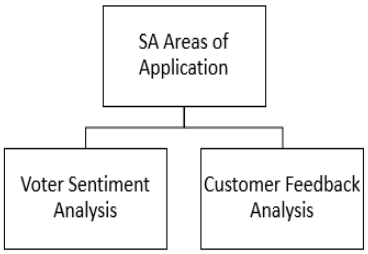


Figure 1 SA areas of Application

Used sentiment analysis to tudy customer reviews and the relationship between customer and quality of service.

Sentiment and opinion have an important characteristic, namely, they are subjective. Different people have different opinions so this is subjective unlike factual information [3]. Sentiment analysis includes different levels, document level, sentence level, and aspect level [4]. The three main classification levels in SA (Figure 2): document-level, sentence-level, and aspect-level SA. Document-level SA aims to classify an opinion document as expressing a positive or negative opinion or sentiment. As a Sentiment Analysis task, ED can be implemented using ML approach or Lexicon-based approach, but Lexicon-based approach is more frequently used. [1]

Sentiment analysis is the practice of applying natural language processing (NLP) and text analysis

techniques to identify and extract subjective information from text parts.

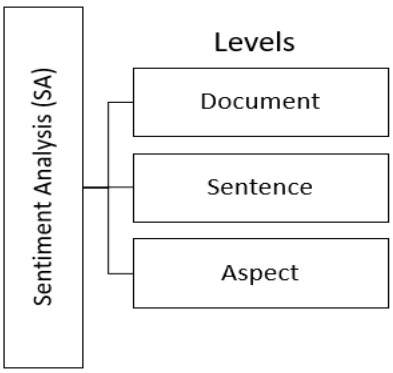


Figure 2 Sentiment Analysis – Levels

Aspect-based sentiment analysis (ABSA) contains three subtasks, namely aspect term extraction, opinion term extraction and aspect-level sentiment classification. In order to make full use of the relationship between the three subtasks, some recent studies have successfully tried to use a unified framework to solve the problem of aspect-based sentiment analysis. people talk about entities that have many aspects (attributes) and they have a different opinion about each of the aspects[4].

II. REVIEW OF SENTIMENT ANALYSIS APPROACHES

In Table 1 , we have reviewed some Mahine learning approaches to Customer Review study**.** In [4] , the authors describes the problem of aspect identification as an information extraction problem.

In this paper the researchers have used a tagged corpus to train a sequence classifier such as a Conditional Random Field (CRF). For aspect-level sentiment analysis, the important first step is to identify the aspects and their associated entities present in customer reviews. Aspects can be either explicit or implicit, where the identification of the latter is more difficult. For restaurant reviews, this difficulty is escalated due to the vast number of entities and aspects present in reviews. The problem of implicit aspect identification has been studied for customer reviews in [6,8]. In this research the authors calculated a score of a given aspect.

In another study [9], the authors try to identify relationships between customers' service evaluations embedded in online reviews and customer satisfaction and demonstrate the unique role of price and location.in the restaurant industry. In [10] the authors predict the sentiment of customer reviews based on a subset of the Yelp Open Dataset

Table 1: Literature Review of Machine learning and Natural Language Processing approaches in Sentiment Analysis

Publication Title	Authors	Publication Type	AI/ML Technique	Result
Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes[11]	RA Laksono, KR Sungkono, R Sarno, CS Wahyuni	Conference Publication	Naive Bayes Method, TextBlob Method	The aim of this study is to analyze restaurant customer reviews from the best 10 restaurants from Tripadvisor reviews. The paper compared the results from two algorithms and concluded the Naive Bayes algorithm has better accuracy than TextBlob sentiment analysis.
A Comparative Study of Sentiment Analysis Using SVM and SentiWordNet [12]	Mohammad Fikri, Riyanarto Sarno	Journal	SentiWordNet, a public dictionary and the support vector machine algorithm, as a complex machine-learning algorithm with a maximum accuracy of 76%).	Since the number of sentences in positive, negative and neutral classes is imbalanced, the oversampling method is implemented. For an imbalanced dataset, the rule-based SentiWordNet and SVM algorithm achieve accuracies of 56% and 76%, respectively. However, for the balanced dataset, the rule-based SentiWordNet and SVM algorithm achieve accuracies of 52% and 89%, respectively.
Sentiment Analysis of Review Datasets Using Naïve Bayes‘ and K-Nn Classifier[13]	L Dey, S Chakraborty, A Biswas, B Bose, S Tiwari	Conference	Naïve Bayes‘ and K-Nn Classifier	The authors came to the conclusion that the SVM classifier can be used successfully for analyzing the customer reviews for products.
Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews.[14]	Hanhoon Kang, Seong Joon Yoo & Dongil Han	Journal	Naive-Bayes and Support Vector Machine	The average accuracy decreases when the accuracies of the two classes (positive or negative review) are expressed as an average value. To address this issue the improved Naïve Bayes algorithm

				is proposed. The result of the experiment showed that when this algorithm was used and some unigrams + bigrams was used as the feature; the gap was reduced by 8 times as compared to when Support Vector Machines were used.

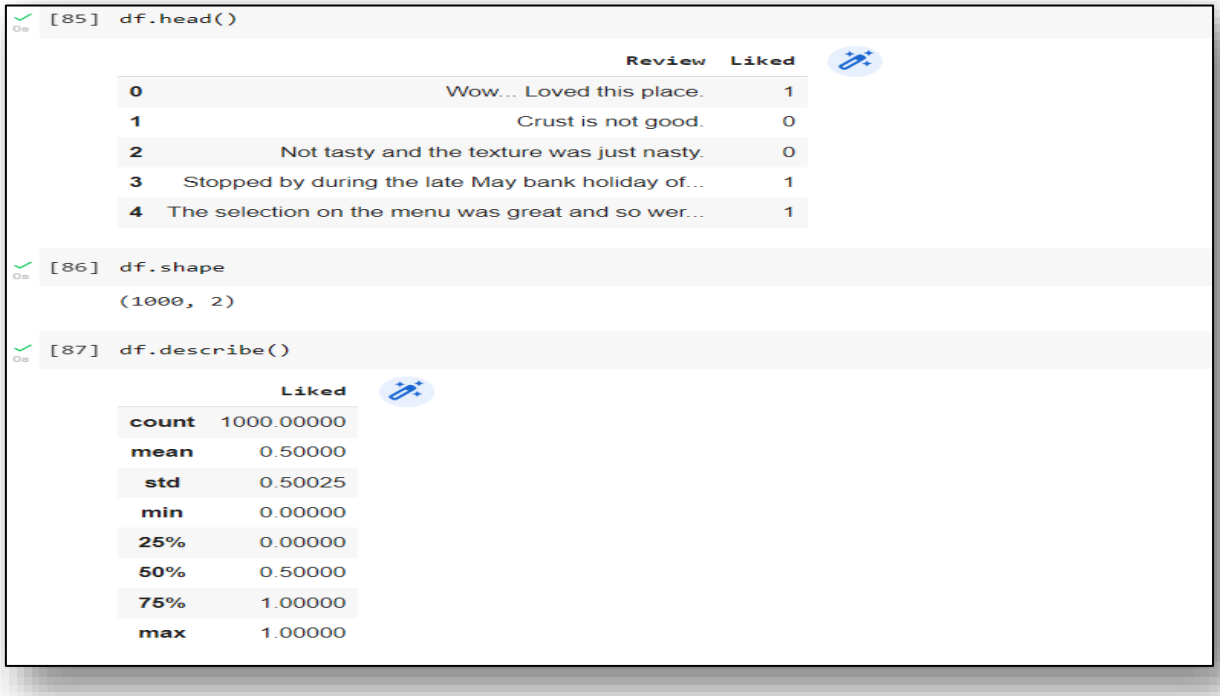
III. METHODOLOGY

3.1. Choosing a dataset

Restaurant reviews Dataset is an open source dataset, which we use for our study. The methodology involves the following steps:

Stage 1 – Exploratory Data Analysis (EDA)

Step 1-Import the python libraries required for EDA into Google Collab file.
Step 2 -Store the restaurant Reviews datasets into a dataframe.
Step 3 –View the statistical information of the dataset.
Step 4 – The dataset consists of two columns, one column is the text of customer review and the second column 'Liked' is the value 0 or 1, which indicates negative or positive sentiment.



```
[85] df.head()
```

	Review	Liked
0	Wow... Loved this place.	1
1	Crust is not good.	0
2	Not tasty and the texture was just nasty.	0
3	Stopped by during the late May bank holiday of...	1
4	The selection on the menu was great and so wer...	1

```
[86] df.shape
(1000, 2)
```

```
[87] df.describe()
```

	Liked
count	1000.00000
mean	0.50000
std	0.50025
min	0.00000
25%	0.00000
50%	0.50000
75%	1.00000
max	1.00000

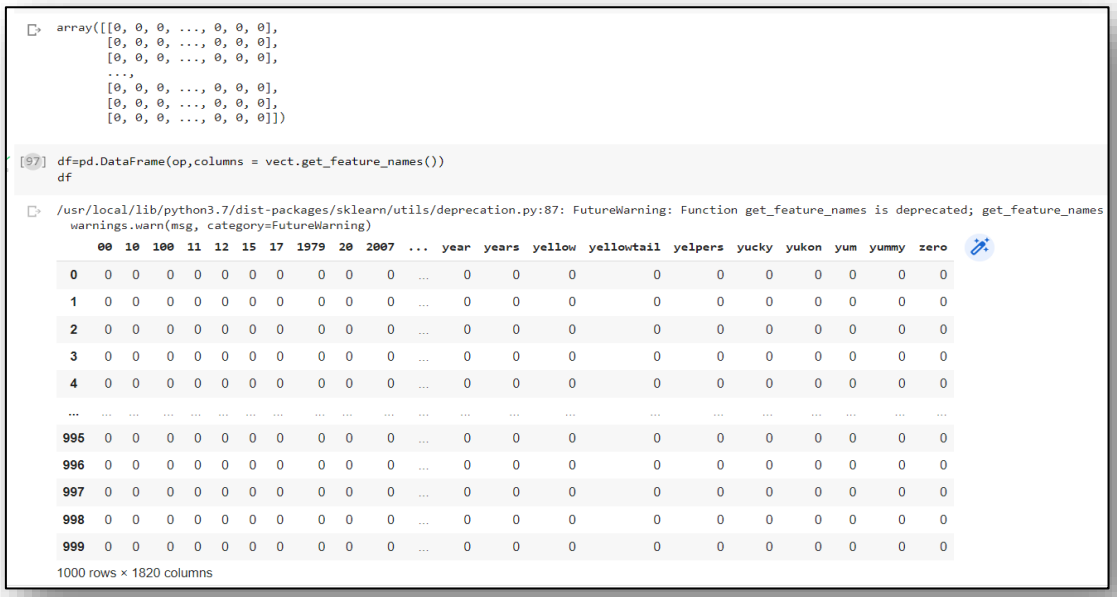
Figure 3 EDA of Customer Review dataset

Stage 2 – Machine learning

Step 1 – Split the dataset into two testing sets and two raining sets using the sklearn.model_selection library.
Step 2- View the four sets.Step 3 – Convert the text to numerical data and store the values into a sparse matrix using CountVectorizer method.
Step 4 – The tokenizing process involves converting all text in the Review column to small case.
Step 5 – The words that are not of significance to form a sentiment are ignored by comparing with a built in list in sklearn library. The CountVectorizer generates a frequency of a word in a particular review.

Stage 3 - Training the Classification model

Step 1 – Using Support Vector Machine (SVC) , fit () method the training sets are passed to the model.
Step 2 – Pass the values of the test dataset to the model to predict the values of the test set.
Step 3 – The accuracy_score method in scikit learn library will help generate the comparative accuracy of our machine learning model. The accuracy of the model is 79%.



```
array([[0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       ...,
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0],
       [0, 0, 0, ..., 0, 0, 0]])
```

```
[97] df=pd.DataFrame(op,columns = vect.get_feature_names())
     df
```

```
/usr/local/lib/python3.7/dist-packages/sklearn/utils/deprecation.py:87: FutureWarning: Function get_feature_names is deprecated; get_feature_names
  warnings.warn(msg, category=FutureWarning)
```

	00	10	100	11	12	15	17	1979	20	2007	...	year	years	yellow	yellowtail	yelpers	yucky	yukon	yum	yummy	zero
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
995	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
996	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
997	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
998	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
999	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

1000 rows × 1820 columns

Figure 4 The CountVectorizer output

Stage 4- Predict the Test results

Step1 –Using the predict() method generate the value and store in y_pred array.

Step 2 – The y_pred array consists of the positive and negative values attained on the training the test set.

```
[122] from sklearn.metrics import accuracy_score
      accuracy_score(y_pred,y_test)

      0.792

[123] import joblib
      joblib.dump(text_model,'positive(1)-negative(0)_review_model')

      ['positive(1)-negative(0)_review_model']

[124] import joblib
      text_model = joblib.load('positive(1)-negative(0)_review_model')

[125] text_model.predict(["Not tasty and the texture was just nasty." ])

      array([0])
```

Figure 5 joblib accuracy score and joblib

Stage 5 – Save the model for predicting the sentiment of new reviews

Step 1 – Using joblib library store the model for future use, (See Fig 5). Step2- Test the new test model and predict the review of the new review as test_model.predict().

Stage 6 – Developing the application using Streamlit Framework

Step 1 - Using the Streamlit library (See Fig. 6) and its methods, we can create an application (See Fig.7) to run the prediction model from a web-browser

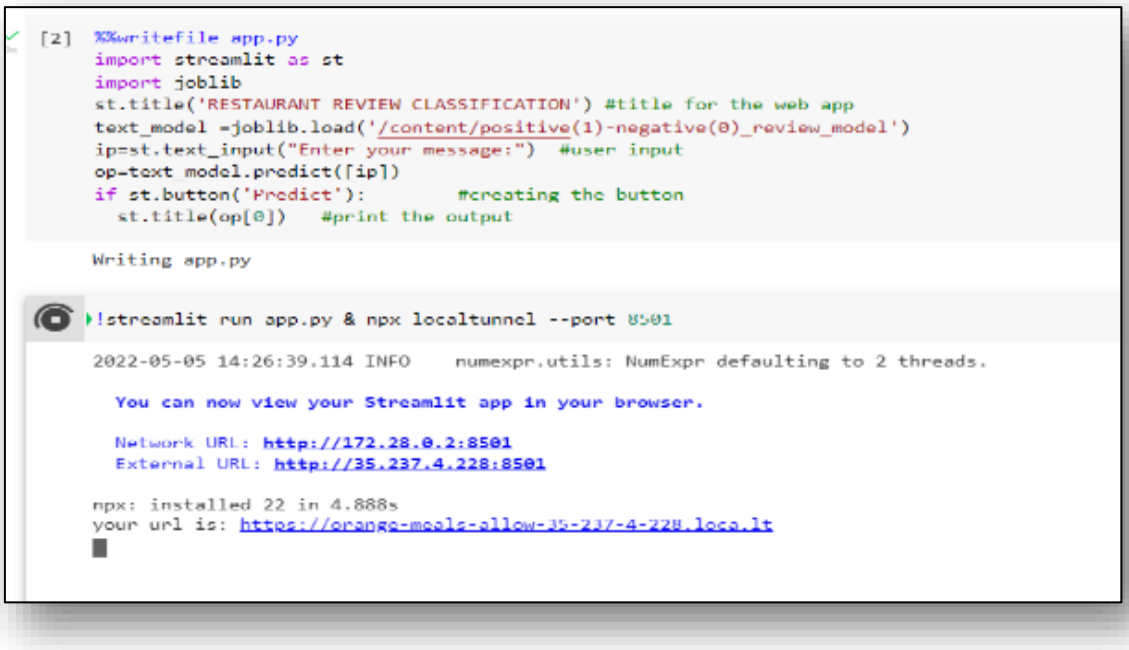


Figure 6 Streamlit Application window

Figure 7 Restaurant Review Classification Output

IV. CONCLUSIONS

I have developed an application for understanding the customer reviews of the restaurant. The customer sentiment can be evaluated using AI/ML techniques and algorithms have learned how to work on support vector classifier and countvectorizer and also I have seen how to use both on the model using pipeline and I have created a model which is able to predict whether the review is positive or negative. I have also seen it using some examples. And I saved the model using the joblib and also retrieved it and used back using the joblib method.

As future work, I wish to develop the dashboard application for this project using Power BI and Tableau. This will help me incorporate more functionality to help the restaurant owners get more information from their customer reviews.

V. ACKNOWLEDGMENTS

I wish to thank my supervisor for providing the motivation to select sentiment analysis as my project work.

REFERENCES

1. Liu, Bing, and B. Liu. "The problem of sentiment analysis." Sentiment Analysis (2015): 16-46.
2. Medhat, Walaa, Ahmed Hassan, and Hoda Korashy. "Sentiment analysis algorithms and applications: A survey." *Ain Shams engineering journal* 5.4 (2014): 1093-1113.
3. Hussein, Doaa Mohey El-Din Mohamed. "A survey on sentiment analysis challenges." Journal of King Saud University-Engineering Sciences 30.4 (2018): 330-338.
4. Feldman, Ronen. "Techniques and applications for sentiment analysis." *Communications of the ACM* 56.4 (2013): 82-89.
5. Bayer, Markus, et al. "Data augmentation in natural language processing: a novel text generation approach for long and short text classifiers." *International Journal of Machine Learning and Cybernetics* (2022): 1-16.
6. Panchendrarajan, Rrubaa, et al. "Implicit aspect detection in restaurant reviews using cooccurence of words." Proceedings of the 7th Workshop on computational approaches to subjectivity, sentiment and social media analysis. 2016.
7. Naumzik, Christof, Stefan Feuerriegel, and Markus Weinmann. "I will survive: Predicting business failures from customer ratings." *Marketing Science* 41.1 (2022): 188-207.
8. Nguyen, T. H., Shirai, K. (2015). Aspect-Based Sentiment Analysis Using Tree Kernel Based Relation Extraction. Computational Linguistics and Intelligent Text Processing, 114-125. https://doi.org/10.1007/978-3-319-18117-2_9
9. Kim, Jaewook, et al. "Why am I satisfied? See my reviews–Price and location matter in the restaurant industry." International Journal of Hospitality Management 101 (2022): 103111.
10. Guda, Bhanu Prakash Reddy, Mashrin Srivastava, and Deep Karkhanis. "Sentiment Analysis: Predicting Yelp Scores." arXiv preprint arXiv:2201.07999 (2022).
11. Laksono, Rachmawan Adi, et al. "Sentiment analysis of restaurant customer reviews on TripAdvisor using Naïve Bayes." 2019 12th International Conference on Information & Communication Technology and System (ICTS). IEEE, 2019.
12. Fikri, Mohammad, and Riyanarto Sarno. "A comparative study of sentiment analysis using SVM and SentiWordNet." Indones. J. Electr. Eng. Comput. Sci. 13.3 (2019): 902-909.
13. Dey, Lopamudra, et al. "Sentiment analysis of review datasets using naive bayes and k-nn classifier." arXiv preprint arXiv:1610.09982 (2016).
14. Kang, Hanhoon, Seong Joon Yoo, and Dongil Han. "Senti-lexicon and improved Naïve Bayes algorithms for sentiment analysis of restaurant reviews." Expert Systems with Applications 39.5 (2012): 6000-6010.