HOTEL'S CUSTOMERS FEEDBACK ANALYSIS USING SENTIMENT ANALYSIS

Billy Gatete[1,2] and Dr. Papias Niyigena[1]

University of Lay Adventist of Kigali, Rwanda

ABSTRACT

In hotel businesses, sentiment analysis allows companies to accurately identify negative or positive sentiment about their products or services and take the necessary steps to address these areas. Customers prefer to book a hotel with a large number of positive reviews. There are various sources to find reviews about the hotel to better understand its reputation. Therefore, customer reviews play an important role for business owners to improve their services. This study uses Logistic Regression classifier to analyze and predict customers' polarity in sentiment analysis.

Keywords: Hotel, Customer feedback, customers' polarity, sentiment analysis, machine learning

INTRODUCTION

In recent years, consumers have changed the way they give feedback. In short, the purpose of accessing and using the websites or social media has become more interactive as people seek information about their daily life experiences and the exchange of products and services they consume(Franciele et al., 2021).

Customers believe that the best way to find a good restaurant is to ask a stranger. It's a problem for him to make a decision if the customer doesn't let anyone ask him. Hotel managers ask different questions like: what makes a good restaurant? What are the major concerns of customers for a great meal? Common sense may give general answers such as good food, great service, and a comfortable environment, but it may not apply to different types of restaurants(Yu et al., 2017).

Opinion mining plays a very important role in the decision making of all customers. If a customer does not receive information from a restaurant customer, they suddenly go to an online

website with more information about the restaurant. Sentiment analysis is called finding opinions from a large amount of data that can help you analyze which restaurant is best for customers who have direct access to good restaurant reviews(C et al., 2019).

Today, sentiment analysis is becoming a great and excellent theme for development and is useful from many applications. The amount of information that can be collected from the Internet is constantly increasing rapidly(C et al., 2019). Sentiment analysis systems help transform unstructured information into structured information about public reviews, products, services, and brands. This is useful in commercial areas such as marketing analysis, open transactions, product reviews, promoters and ratings, product feedback, and product service.

In this paper, the main aim is to analyze customers' feedback, predict customers' polarity, score, classify and visualize the feelings and opinions that your customers display in their reviews of hotel using Logistic Regression machine learning.

RELATED WORKS

Various studies have been conducted on sentiment analysis and machine learning concepts to determine the customers' polarity from customers' feedback. Below are some of these research scientists reviewed in this study.

According to (Kasper & Vela, 2011), Travel planning and hotel bookings on the website have become important commercial uses. Internet sharing has become an important tool for expressing customer thinking about a particular product or service. In this study, Multinomial Naive Bayes classifier was used as a supervised probabilistic learning method that handles the number of occurrences of each word in a document and support vector machine (SVM) that uses a linear kernel.

In the (Lai & Raheem, 2020) study, An unsupervised approach, called lexicon-based method of sentiment analysis, is commonly used to infer the emotions represented by a word / lexicon by showing polarity and subjectivity was used. The study focused on online reviews commonly shared by hotel guests and reflects their positive, negative, neutral and other mood experiences and satisfaction. Sentiment analysis performed using a lexicon-based approach has high accuracy but low recall, resulting in poor performance.

According to (Claire et al., 2018), Most review sites provide both textual and numerical hotel ratings, and both sources are commonly used by researchers to express customer sentiment and opinion. However, opinion is a difficult concept to measure, so it is clear whether it is so,

depending on the relationship between these two sources. Researchers used an entropy-based support vector machine to classify the positive and negative emotions of hotel reviews from the website Ctrip and calculate the ratio of positive and negative emotions for each review.

As a finding from previous studies, all studies outcomes reviewed above lead to classify the positive and negative emotions of hotel reviews using different machine learning. While these models show good results, they show weaknesses in modeling the key contextual functions required for sentiment analysis especially lexicon based machine learning.

METHODOLOGY

This proposed work is expected to be designed for text analysis using restaurant reviews datasets given by the customers for a food in restaurant. Using target attribute values helps to classify textual data using a Logistic Regression classifier. The Figure 1 shows the architecture diagram of Logistic Regression classifier for predicting the positive, negative or neutral of sentiment analysis.
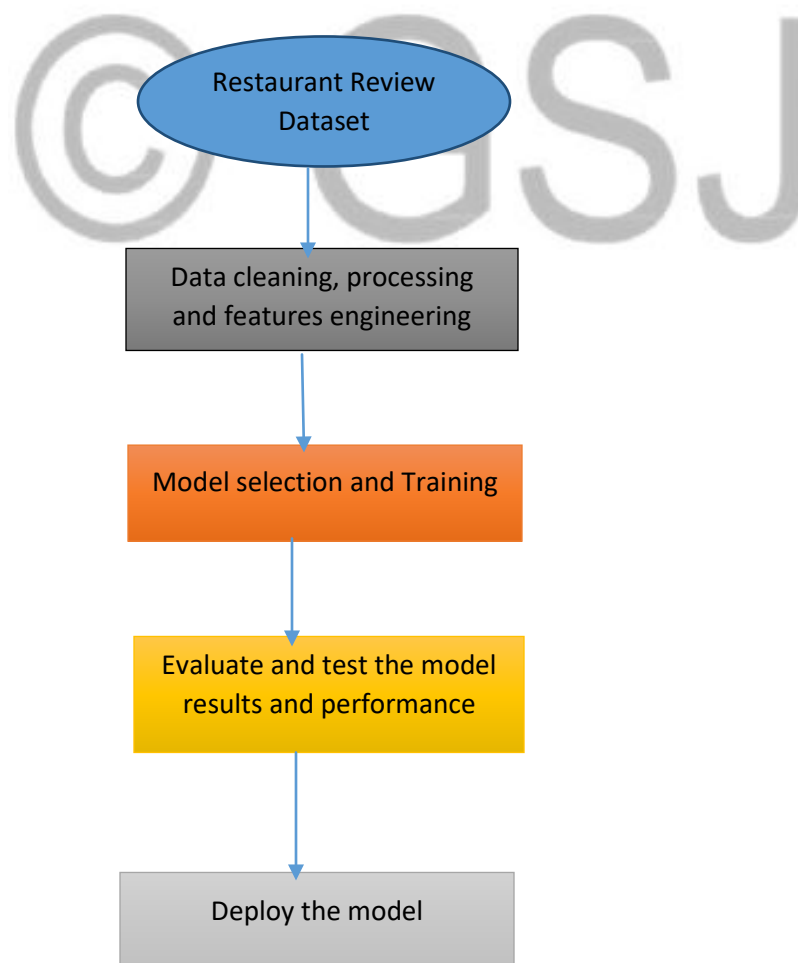

Figure 1: The architecture diagram of Logistic Regression classifier

In this paper, only secondary data was collected. The data were downloaded from the **kaggle.com** dataset, a platform of big data sets. The dataset in csv has 1000 rows and 2 columns that represent Label and Review feedback (in form of texts). Much focus will be put on the data downloaded from kaggle.com whose data will be taken and fit into the machine learning-based sentiment analysis algorithm to meet our specific objectives.

Researcher plotted the graph using Matplotlib, a plot library in the Python programming language, and its numerical math extension Numpy. These graphs help you interpret trends in statistical data. Different performance evaluation metrics were used to evaluate the machine learning. The area under the curve, classification report, accuracy, recall (sensitivity and F1-score were used to evaluate the performance of the learner.

RESULTS AND DISCUSSION/FINDINGS

The result showed that Logistic Regression classifier produced the accuracy of 84%, recall of 84%, F1-score 84 % and the precision of 84% as shown in the table below.

Table 1: Classification report for Logistic Regression

	precision	recall	f1-score	support
0	0.87	0.82	0.85	108
1	0.81	0.86	0.83	92
accuracy			0.84	200
macro avg	0.84	0.84	0.84	200
weighted avg	0.84	0.84	0.84	200

The area under the curve of this classifier was 84.13%. This meant that the classifier was able to distinguish between positive and negative class values. The classifier detected more true

positives and true negatives than false negatives and false positives as shown in the figure below
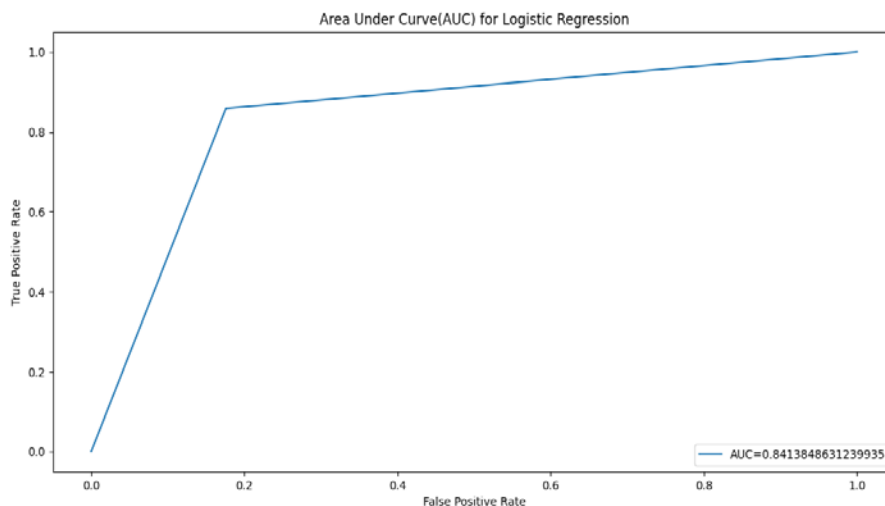

Figure 2: Area under the curve of Logistic Regression

Text analysis using restaurant reviews datasets given by the customers for a food in restaurant shown that the reviews contained 32.81% of negative reviews, 39.37% of the neutral reviews and 27.82% of positive reviews as shown in the following diagram. More comments were neutral followed by negative. The reviews contained few positive reviews. This meant that the department did not give good services.
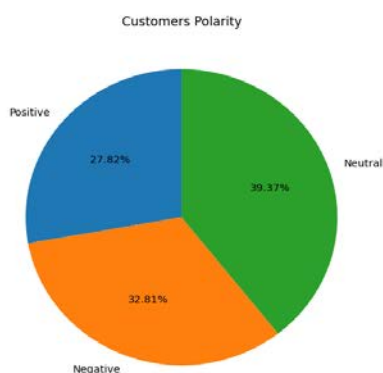

Figure 3: Customer's polarity

The word cloud of the dataset is a visual representation of textual data. The result showed that "food", "place", "good", "service", "great" and more other words appeared the most as shown below.

Figure 4: Word cloud of customers' feedback

A 3 gram (trigram) which is a sequence of three words such as "the food good", "would recommend place", "the food terrible" and so on was plotted with a 2 gram (bigram) that composed with two words in the dataset. The results from customers' feedback shown that "this place", "the food" appeared the most as shown below

```
Words head :

                     term       rank
3467        the food good   3.687271
4176  would recommend place 2.000000
3472     the food terrible  2.000000
4118        wo going back   1.856351
4171        would go back   1.555110
241         back anytime soon 1.403865
1397      going back anytime 1.403865
```

Figure 5: Trigram of customers' feedback

```
Words :
            term        rank
4137    this place   8.799255
3975     the food    7.357092
1614      go back    6.862445
4033    the service  5.004646
1425     food good   4.257144
1741    great food   3.535416
4617      wo back    3.206378
```

Figure 6: Bigram of customers' feedback

CONCLUSION

This study intended to analyze and predict the customers' polarity from the feedback downloaded from Kaggle.com. It visualized the polarities that can help the managers and the owners or investors to take the right decision. Logistic Regression classifier was used, it produced good accuracy, recall, f1-score, precision and area under the curve. The results showed that 32.81% of the feedback are negative, 39.37% of the feedback are neutral and 27.82% of the feedback are positive reviews. To better visualize data, predict polarity and analyze data, different python libraries were used such as Matplotlib, Numpy, sklearn, pandas, etc... Word cloud, bigram and trigram highlighted the common used words or set or words as shown with the graphs. One further direction is to use advanced machine learning algorithms or other advanced techniques that produce better results especially on small dataset.

REFERENCES

1. C, S., Ravikumar, D. P., & M.J, M. A. (2019). Sentiment Analysis of Customer Feedback on Restaurant Reviews. *SSRN Electronic Journal*. https://doi.org/10.2139/ssrn.3506637

2. Claire, A. C. E., 野中尋史, & 平岡透. (2018). Relation Analysis between Hotel Review Rating Scores and Sentiment Analysis of Reviews by Chinese Tourists Visiting Japan. 産業応用工学会論文誌, *6*(2), 95–99. https://doi.org/10.12792/jjiiae.6.2.95

3. Franciele, C., Christina, T., Ruiz, D., Manosso, F. C., & Cristina, T. (2021). Using sentiment analysis in tourism research: A systematic, bibliometric, and integrative

review. *Journal of Tourism, Heritage & Services Marketing*, *2021*(2), 16–27. https://doi.org/10.5281/zenodo.5548426ï

4. Kasper, W., & Vela, M. (2011). Sentiment Analysis for Hotel Reviews. *Proceedings of the Computational Linguistics-Applications Conference*, *231527*(March), 45–52. www.tripadvisor.com

5. Lai, S. T., & Raheem, M. (2020). *SENTIMENT ANALYSIS OF ONLINE CUSTOMER REVIEWS FOR HOTEL INDUSTRY : AN APPRAISAL OF HYBRID APPROACH*. 1355–1359.

   https://www.researchgate.net/publication/347565122_SENTIMENT_ANALYSIS_OF_ONLINE_CUSTOMER_REVIEWS_FOR_HOTEL_INDUSTRY_AN_APPRAISAL_OF_HYBRID_APPROACH

6. Yu, B., Zhou, J., Zhang, Y., & Cao, Y. (2017). *Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews*. http://arxiv.org/abs/1709.08698