Original Article
ISSN (Online): 2456-8651

International Journal of Engineering Science Technologies
January-February 2022 6(1), 41–50

APPLYING OPINION MINING ALGORITHMS TO ANALYZE THE USER SENTIMENTS FROM USER REVIEWS TO RATE THE RESTAURANTS IN SULTANATE OF OMAN

Christine Mariam Joseph [1], Arunima T.A [1], Vinu Sherimon [2], Khoula Mohamed Mubarak Ali Al Amri [2], Nidhal Ali Salim Al Amri [2]

[1] Department of Computer Science and Engineering, Saintgits College of Engineering, Kottayam, Kerala, India
[2] Department of Information Technology, University of Technology and Applied Sciences, Muscat, Sultanate of Oman.

ABSTRACT

Humans now spend the bulk of their time on the Internet, thanks to the huge rise of social media and other applications. Organizations and individuals have a shared platform to express their thoughts and ideas on any entity in the globe. Every website has a massive amount of information about products and services. Through social media tools, we, as humans, may remark on any subject or incident. Most of the time, these public opinions are considered by organizations and individuals to better corporate operations and strategies, or in the decision-making process. Although these user opinions are practical in decision-making processes, manually analyzing and summarizing them is a difficult effort. It is a lengthy process. As a result, having automated opinion mining techniques to analyze user sentiments is critical. Since the beginning of social media, sentiment analysis has been a hot research area. Many studies have been conducted in various fields. The tourism sector is highlighted as one of the important topics for diversification in Oman Vision 2040. This study suggests an analysis of user reviews of Omani restaurants, categorizing them as positive or bad, and ranking the restaurants according to the reviews. The reviews are analyzed using opinion mining methods.

Keywords: Hotels, social media, Reviews

Received 1 January 2022
Accepted 15 January 2022
Published 9 February 2022

Corresponding Author
Vinu Sherimon,
vinusheri@hct.edu.om

DOI 10.29121/IJOEST.v6.i1.2022.263

Funding: The research leading to these results has received funding from the Research Council (TRC) of the Sultanate of Oman under the Block Funding Program BFP/URG/ICT/20/030.

1. INTRODUCTION

Every activity in which humans participate has its own set of opinions, which are intertwined in our behavior. Whether we realize it or not, we constantly offer our opinions and thoughts wherever they are required. When it comes to making decisions, we also seek the advice of others to some level. We have a massive amount of data stored in several places, including online forums, social media, and blogs Liu Bing (2012). They contain people's thoughts about events, products, and services, among other things. When we want to buy an eBook, for example, we normally look at the user reviews to see what others have to say. Positive, negative, or neutral feedback is possible. However, these reviews assist us in deciding. The study of mechanically analyzing people's emotions, sentiments, opinions, and perspectives about any entity in the world is known as sentiment analysis or opinion mining Liu Bing (2012).

Oman has a diverse range of eateries, both domestic and international.

We are continuously on the lookout for good, flavourful, and high-quality cuisine cuisine. As a result, a system that allows user comments and reviews on restaurants and provides automatic ranking of restaurants would be quite valuable. Opinion mining methods are used in the proposed study to assess user reviews of eateries in Oman. All descriptions of restaurants in Oman will be examined as part of the proposed study. Users' feedback is considered in the analysis. The feedback will be classified as favourable or negative using supervised machine learning methods. These algorithms are mostly used to solve categorization issues. In a random forest method, for example, many trees are produced, and the predictions of each tree are examined before voting on the best answer.

The main objective of this research is to analyse the user feedback, reviews, and comments about restaurants in Oman using opinion mining algorithms. Conduct literature review on various restaurant review rating systems based on data mining methodologies to examine the advantages and disadvantages of such systems Collect the details about different restaurants in Oman.

The rest of the paper is organized as follows: Related work is presented in Section 2. Section 3 describes the methodology of the research which includes dataset details, techniques used for building a Recommendation System, architecture of the proposed application, and K-means clustering algorithm. Experimental results are given in Section 4 and Conclusion is given in Section 5. Section 6 includes Future work followed by References.

2. RELATED WORK

Prior to the year 2000, there was no sentiment analysis research since user reviews were not available in digital form Liu Bing (2012). Users didn't have many options to share their opinions back then because social media sites didn't exist. Sentiment analysis is now at the forefront of research, because of the phenomenal expansion of social media applications.

A recent Romanian study Ruseti et al. (2020) focused on extracting 200,000 game reviews and categorizing them as good, negative, or neutral. The study used three different algorithms: support vector machines, multinomial Naive Bayes, and deep neural networks.

Researchers in Spain developed a cloud-based solution based on Python for analysing the feelings expressed in restaurant reviews in the province of Granada Agüero-Torales et al. (2019). The software examines TripAdvisor.com reviews as well as the province's top 10 restaurants. The number of reviews and the user rating are mentioned Agüero-Torales et al. (2019).

Naïve Bayes algorithm was used in Laksono et al. (2019) to analyse the sentiments of customer reviews about Surabaya restaurant in Indonesia on TripAdvisor.com. The results of the research show that Naïve Bayes approach is better than Text Blob method with an accuracy of 72.06% Laksono et al. (2019).

In Jonathan et al. (2019) a Random Forest Classifier was used to predict the feelings of restaurant reviews from Zomato, a restaurant rating app. The data was obtained from Kaggle and analysed using the Python Scikit-Learn module to predict sentiments Jonathan et al. (2019). The investigation is said to have a 92 percent accuracy rate Jonathan et al. (2019).

Another fascinating study Nakayama et al. (2018) looked at if there was a cultural difference in Yelp evaluations of ten well-known Japanese foods. In the study, two types of reviewers were chosen: Western reviewers and Japanese reviewers. The findings of the study reveal that reviewers with a Western background use more thorough reviews to rate the restaurant features such as service quality and food quality Nakayama et al. (2018).

Sentiment analysis was used by academics in the United States in 2017 to discover the characteristics of various eateries Yu et al. (2017). The feelings were deduced from the word frequency of user evaluations using a Support Vector Machine (SVM) model Yu et al. (2017). Based on these, a polarity index was calculated, and specific restaurant attributes were identified.

3. METHODOLOGY

3.1. LIBRARIES

Here are some of the most important libraries we'll be using.

- Natural Language Toolkit (NLTK): the most well-known Python suit of libraries for Neural network models
- Gensim: it is a toolbox for subject modelling and vector space modelling
- Scikit-learn: an important Python library that encompasses a set of useful machine learning and statistical modelling techniques.

NLTK

The Natural Language Toolkit, or NLTK for short, is a Python-based collection of tools and programmes for conceptual and statistical natural language processing for English. Graphical demos and sample data are included in NLTK. It comes with a cookbook and a book that describes the basic concepts behind the language processing jobs that the toolkit supports.

Gensim

Gensim is an open-source package that uses contemporary statistical machine learning to perform unsupervised topic modelling and natural language processing. For performance, Python and Cython is used to write Gensim. Unlike many other machine learning software packages, Gensim is designed to handle large text datasets using data streaming and progressive online techniques rather than in-memory computation.

Scikit-learn

Scikit-learn is a free and significant library in Python machine learning package. Agüero-Torales et al. (2019) It is meant to interact with the Python mathematical and scientific libraries NumPy and SciPy, and features support vector machines, gradient boosting, k-means, random forests etc.

3.2. DATA COLLECTION TOOLS

The data collection methods that come under descriptive research include surveys. When we want to collect a small amount of information from many people, we employ surveys. Online surveys are a more effective and simple way to gather data. So, the main tool that is used to conduct this research is online surveys. It is very easy to use and saves us time.

We'll take data from hotel reviews in this case. Each observation is made up of a single hotel's customer evaluation. Each customer review contains both a narrative description of the client's hotel feels and an overall rating.

We want to know whether each textual review correlates to a positive (satisfied) or negative (dissatisfied) review. The overall grade of the reviews might range from 2.5 to 10 out of ten. To make the problem easier to understand, we'll divide them into two groups:

- poor reviews having an overall rating of less than five stars
- good reviews having an overall rating of more than or equal to five stars

The problem here is that only the pure textual data from the review can be used to forecast this information.

A custom dataset created by surveying different online hotel review websites in Oman are used along with 515k hotel ratings data in Europe dataset. Booking.com provided the information. Everyone already has access to all the info in the file. Booking.com is the original owner of the data. The dataset is large and informative.

3.3. DATA PROCESSING

To begin, we must first load the raw data. Each literary review is divided into two parts: a favourable and a negative section. We put them together so we could begin with only basic text data and no extra information.

	review	is_bad_review
0	I am so angry that i made this post available...	1
1	No Negative No real complaints the hotel was g...	0
2	Rooms are nice but for elderly a bit difficul...	0
3	My room was dirty and I was afraid to walk ba...	1
4	You When I booked with your company on line y...	0

Figure 1 Data set review are divided into positive and negative

To expediate the calculations, review data is sampled. If a user does not write a negative feedback comment, our database will show "No Negative." Only positive comments were recorded with the value "No Positive". Those sections must be removed from our texts.

The next stage is to clean the text data using a variety of operations. We use a custom 'clean text' function to clean textual data, which performs many conversions:

- Lowercase the text characters
- Eliminate the punctuation and tokenize the content (divide it into words/parts).
- Delete any terms that include numbers that aren't necessary.
- Remove unnecessary stop words like 'the,' 'a,' and 'this,' among others.
- Utilize the WordNet lexical database, add a tag to each word to indicate whether it is a noun, a verb, or something else (Part-of-Speech (POS) tagging.
- Convert each word into its base form (e.g., cats → cat) which is the process of lemmatizing.

3.4. FEATURE ENGINEERING

We start with sentiment analysis since we feel that the reviews of the customer are closely tied to how customers experienced throughout their hotel visit. We use Vader, which is a sentiment analysis component of the NLTK module. Vader consults a dictionary of words to determine which are positives and which are negatives. It also considers the context of the statements while calculating sentiment scores. For each word, Vader returns four values: positive value, negative value, neutral value, and an aggregate score which is the summary of the all the other scores.

These four values will be used as features in our dataset. Next, we add some simple metrics for every text:

- number of characters in the text
- number of words in the text

For each review, the next step is to extract vector representations. The Gensim module generates a numerical vector representation of each word in the corpus (Word2Vec) based on the circumstances in which they occur. Shallow neural networks are used to do this. What's intriguing is that representation vectors for comparable words will be similar.

Any text can be converted into numerical vectors (Doc2Vec) using the term vectors. The interpretations of the same texts will be identical, which is why those vectors can be utilized as training variables.

We must first feed our text data into a Doc2Vec model to train it. We may obtain such representation vectors by using this model on our reviews.

Finally, we include the Term Frequency - Inverse Document Frequency (TF-IDF) values for each phrase and document.

We can just count the number of times each word exists in each document. But this technique has the drawback of ignoring the relative importance of words in the texts. A word that occurs in almost every text is unlikely to yield useful information for analysis. On the other hand, words that are not common, may have a broader meaning range.

The TF-IDF measure tackles this difficulty as follows:

- TF estimates the standard number of times a word occurs in a text.
- IDF estimates the relative value of this word depending on how many texts it exists in.

We include TF-IDF columns for every phrase that occurs in at least 10 different texts to filter some of the terms and reduce the size of the result.

	review	is_bad_review	review_clean	neg	neu	pos
488440	Would have appreciated a shop in the hotel th...	0	would appreciate shop hotel sell drinking wate...	0.049	0.617	0.334
274649	No tissue paper box was present at the room	0	tissue paper box present room	0.216	0.784	0.000
374688	Pillows Nice welcoming and service	0	pillow nice welcome service	0.000	0.345	0.655
404352	Everything including the nice upgrade The Hot...	0	everything include nice upgrade hotel revamp s...	0.000	0.621	0.379
451596	Lovely hotel v welcoming staff	0	lovely hotel welcome staff	0.000	0.230	0.770

Figure 2 Cleaned data

3.5. EXPLORATORY DATA ANALYSIS

Let's take a closer look at our data to gain a better understanding of it:

```
[ ] # show is_bad_review distribution
    reviews_df["is_bad_review"].value_counts(normalize = True)

    0    0.956761
    1    0.043239
    Name: is_bad_review, dtype: float64
```

Figure 3 Positive and Negative review distribution

Because less than 5% of our reviews are unfavourable, our dataset is significantly skewed. This knowledge will come in handy during the modelling process. The majority of the words are hotel-related, such as room, breakfast, staff etc. Perfect, admired, pricey, dislike, and other phrases are more applicable to the consumer's hotel stay experience.

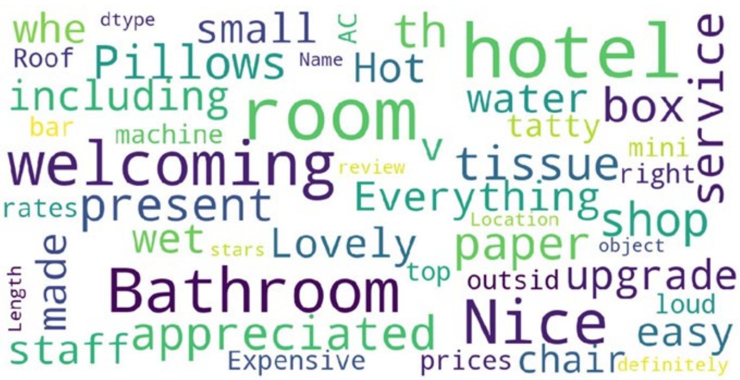

Figure 4 Word Cloud

review	pos
A perfect location comfortable great value	0.931
Clean comfortable lovely staff	0.907
Friendly welcome Comfortable room	0.905
Good location great value	0.904
Clean friendly and comfortable	0.902
Good value amazing location	0.901
breakfast excellent Clean comfort	0.899
Great place I enjoyed	0.881
Beautiful Quirky Comfortable	0.878
Lovely comfortable rooms	0.877

Figure 5 Positive Reviews

review	neg
No dislikes LOCATION	0.831
A disaster Nothing	0.804
A bit noisy No	0.796
Dirty hotel Smells bad	0.762
Very bad service No	0.758
Window blind was broken	0.744
no bad experience location	0.740
nothing great clean comfortable quite hotel	0.733
It was awful No	0.722
Very bad atmosphear noisy weird smells unfrie...	0.713

Figure 6 Negative Reviews

Among the most scathing reviews, there are a few mistakes: Vader misinterprets 'no' and 'nothing' as negative words when they are intended to express that the hotel had no problems. Fortunately, most of the reviews are negative.

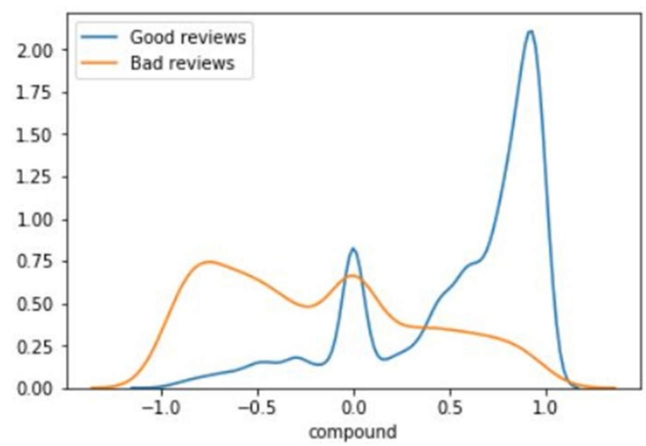

Figure 7 Distribution of the review's sentiments

The graph above depicts the distribution of review feelings between positive and negative reviews. We can see that Vader considers most of the positive evaluations to be quite positive. On the other side, negative evaluations have lower compound sentiment value. This means that the computed sentiment features in the past will be important in our modelling procedure.

3.6. MODELLING REVIEWER_SCORE

We start by deciding the features we'll use to prepare our model. After that, we divided our data into two components:

- A component to train our model

- A component to assess the performance of the model

For our predictions, we'll utilize a Random Forest (RF) classifier.

	feature	importance
3	compound	0.040894
2	pos	0.024434
0	neg	0.023542

Figure 8 Feature importance

The most important characteristics are those derived from the preceding sentiment analysis. In our training, the vector interpretations of texts are also quite important. Some words appear to have a decent amount of significance

4. RESULTS
4.1. ROC CURVE

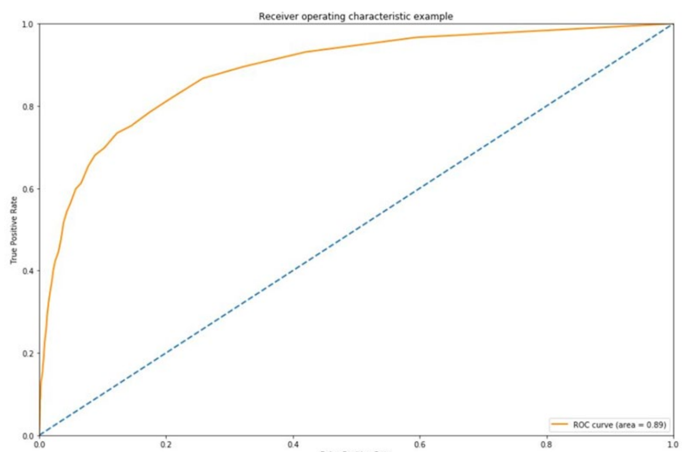

Figure 9 ROC curve

The Receiver Operating Characteristic (ROC) curve is a useful graph for summarizing the effectiveness of our classifier. The higher the curve is above the diagonal line, the better the forecasts are. But we should not utilize the area under the ROC curve to judge the quality of our model, although it is good.

FPR (False Positive Rate) = # False Positives / # Negatives is the formula that correlates to the x axis of the ROC curve.

Because our dataset is unbalanced, the # Negatives correlate to our number of positive reviews, which is unusually high. This indicates that, even if any False Positives occur, our FPR will remain low. Our model will be able to predict a large number of false positives while keeping the false positive rate low, hence raising the true positive rate and artificially enhancing the AUC ROC metric.

4.2. PC CURVE

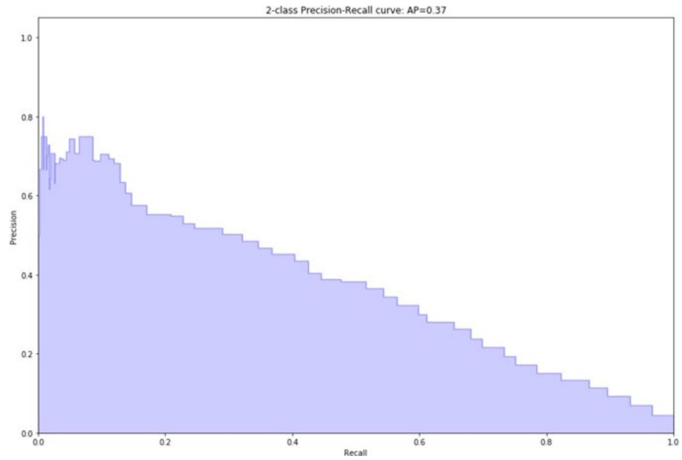

Figure 10 PC curve

In this unbalanced circumstance, Area Under the Curve Precision Recall (AUC PR), also known as AP, is a better statistic (Average Precision).

When we raise the recall, we can see that the precision diminishes. This demonstrates the importance of selecting a prediction threshold that is tailored to our requirements. If we want a high recall, we should select a low prediction threshold, which allows us to find many positive class observations with a low degree of precision. On the other side, if we want to be very confident in our predictions but don't mind missing a few good ones, we should select a high threshold that gives us high precision but low recall.

The AP measure can be used to determine whether our model outperforms another classifier. We may compare the quality of our model to a basic decision baseline to see how well it works. To achieve it, assume a random classifier that predicts the label 1 for half of the time and predicts 0 for the next half.

The precision of such a classification model would be 4.3 percent, which relates to the fraction of positive values. The precision would remain constant for each recall value, resulting in an AP of 0.043. Our model's AP is roughly 0.35, which is more than 8 times greater than the random method's AP. This indicates that our model has a high level of predictability.

It is entirely possible to make predictions using simply raw text as input. The most crucial aspect is being able to obtain appropriate information from this basic data source.

5. CONCLUSION

This research proposes analysing user reviews of Omani restaurants, classifying them as favourable or negative, and ranking the restaurants based on the reviews. Opinion mining techniques are used to examine the reviews. We have developed a model for this. The average precision of the presented model is around 0.35, which is eight times greater than the average precision of the random technique. This demonstrates that our model is extremely predictable. In future, we are planning to develop analysis models to check the sentiments of tweets in Twitter.

REFERENCES

Agüero-Torales, M. M., et al. (2019) "A cloud-based tool for sentiment analysis in reviews about restaurants on TripAdvisor." Procedia Computer Science 162 : 392-399. Retrieved from https://doi.org/10.1016/j.procs.2019.12.002

Jonathan, Bern, Jay Idoan Sihotang, and Stanley Martin. (2019) "Sentiment Analysis of Customer Reviews in Zomato Bangalore Restaurants Using Random Forest Classifier." Abstract Proceedings International Scholars Conference. Vol. 7. No. 1. Retrieved from https://doi.org/10.35974/isc.v7i1.1003

Laksono, Rachmawan Adi, et al. (2019) "Sentiment Analysis of Restaurant Customer Reviews on TripAdvisor using Naïve Bayes." 2019 12th International Conference on Information & Communication Technology and System (ICTS). IEEE, Retrieved from https://doi.org/10.1109/ICTS.2019.8850982

Liu, Bing. (2012) "Sentiment analysis and opinion mining." Synthesis lectures on human language technologies 5.1 : 1-167. Retrieved from https://doi.org/10.2200/S00416ED1V01Y201204HLT016

Nakayama, Makoto, and Yun Wan. (2018) "Is culture of origin associated with more expressions ? An analysis of Yelp reviews on Japanese restaurants." Tourism Management 66 : 329-338. Retrieved from https://doi.org/10.1016/j.tourman.2017.10.019

Oman Vision (2040). https://www.2040.om/en/#Oman2040 (Accessed on Jan 12,2020) 9. Oman National Tourism Strategy. Retrieved from https://omantourism.gov.om/wps/wcm/connect/mot

Ruseti, S., Sirbu, M. D., Calin, M. A., Dascalu, M., Trausan-Matu, S., & Militaru, G. (2020). Comprehensive Exploration of Game Reviews Extraction and Opinion Mining Using NLP Techniques. In Fourth International Congress on Information and Communication Technology (pp. 323-331). Springer, Singapore. Retrieved from https://doi.org/10.1007/978-981-15-0637-6_27

Yu, Boya, et al. (2017) "Identifying Restaurant Features via Sentiment Analysis on Yelp Reviews." arXiv preprint arXiv : 1709.08698. Retrieved from https://arxiv.org/abs/1709.08698