INDIAN JOURNAL OF SCIENCE AND TECHNOLOGY

RESEARCH ARTICLE

Sentiment Analysis and Customer Satisfaction Factors Based on LSTM and Topic Modeling

Sakshi Nazirkar[1]*, **Siddhivinayak Kulkarni**[1]

1 School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Pune, 411038, Maharashtra, India

OPEN ACCESS

Received: 12-05-2023

Accepted: 19-06-2023

Published: 22-07-2023

Citation: Nazirkar S, Kulkarni S (2023) Sentiment Analysis and Customer Satisfaction Factors Based on LSTM and Topic Modeling. Indian Journal of Science and Technology 16(28): 2126-2132. https://doi.org/10.17485/IJST/v16i28.1109

* **Corresponding author**.

sakshi.nazirkar.9@gmail.com

Funding: None

Competing Interests: None

Published By Indian Society for Education and Environment (iSee)

ISSN
Print: 0974-6846
Electronic: 0974-5645

Abstract

Objective: To predict sentiment of the Airbnb text reviews using Long Short Term Memory (LSTM). To improve the accuracy and performance metrics. To identify customer satisfaction and dissatisfaction factors of the Airbnb customers using Sentiment Analysis and Topic Modeling. **Method**: The study is divided into two parts after performing necessary pre-processing steps. First part focuses on sentiment analysis using LSTM. Dataset is created by combining review data of 3 cities, then, operations like pre-processing, sentiment analysis, label column creation, under sampling etc. were conducted. After this, data was trained on the configured LSTM Model. The second part of the study was Topic Modeling after applying Sentiment Analysis, on an Airbnb dataset, to derive and understand customer satisfaction and dissatisfaction factors.**Findings**: Sentiment Analysis using LSTM showed training accuracy of 96.37%.and testing accuracy of 93.89%. The performance metrics showed promising results. The topics found for negative and positive sentiment portraying the customer satisfaction and dissatisfaction factors after Topic Modeling align with the existing literature findings and are important to generalize the existing literature as well. **Novelty**: Improved performance metrics like Accuracy, F1-score and Recall for sentiment analysis using LSTM. Results stating customer satisfaction and dissatisfaction factors add value to the existing literature and help to generalize findings.

Keywords: Natural Language Processing; Sentiment Analysis; Topic Modeling; Deep Learning; LSTM; Customer Satisfaction

1 Introduction

The sharing economy can be defined as, a system in which people can share their possessions with others for free or with a fee, with the help of Internet, for e.g., Airbnb. Previously, comparative studies were done for sentiment analysis using different deep learning approaches like Recurrent Neural Network(RNN), Gated Neural Networks, LSTM, Bi-LSTM etc., on datasets like Kaggle datasets, Yelp dataset, Sentiment140 . The results obtained showed 81% and 32% F1-score for positive and negative sentiment on Kaggle dataset[(1)], 95% accuracy on binary sentiment classification using LSTM on Yelp dataset[2], and 82% accuracy using RNN on the Sentiment140 dataset[3]. The result of this study shows a significant improvement in the training and testing accuracy for Sentiment analysis using LSTM approach. As data from different cities is helpful to generalize the results[4], this study obtained data for 3 different cities from InsideAirbnb.com. The approach for sentiment analysis using LSTM, included conditional labelling the dataset using the Valence Aware Dictionary and sentiment Reasoner (VADER) score, instead of manually labelling the dataset or considering pre-defined labeled dataset. The conditional labelling was used to convert the dataset into supervised learning approach to train the configured LSTM Model. Improved results were obtained, with F1-score of 97% and 57% for positive and negative classified sentiment, respectively.

Sentiment analysis studies can be used for designing effective strategies for tourism destinations[5]. Customer satisfaction and dissatisfaction is not justifiable using the six elementary scores which are provided by InsideAirbnb.com, was stated by authors in[6] and[7]. Therefore, variety of approaches were followed to find different topics that influenced guest sentiment. These approaches included manual intervention, and algorithms or combination of tools like, Latent Dirichlet Allocation (LDA) and Mallet toolkit[8], LDA[9], ROST CM6 and LIWC[10], Microsoft Excel and NVivo Plus version12[11], supervised Latent Dirichlet Allocation (sLDA)[12], VADER[13]. Using VADER and LDA for sentiment analysis and topic modeling on the review text data, this study was able to uncover the customer satisfaction and dissatisfaction factors for Singapore. The results obtained showed topics like price, location, host interaction influencing customer satisfaction, which lie in line with previously obtained results. The previous studies by authors[12–16] show that location, host interaction are major influencing factors for the customer. The results obtained in this study support and added value to the existing literature, and also can be considered as taking a step to justify that location and host interaction are important factors irrespective of the place, for the customers.

2 Methodology

This section is divided into two parts with the same pre-processing steps, first part being Sentiment analysis using LSTM, for positive and negative sentiment. The other part, Topic Modeling on reviews for positive and negative sentiment to understand customer satisfaction and dissatisfaction factors. VADER lexicon is used for obtaining sentiment scores on the cleaned processed text reviews. After sentiment analysis LDA is used for topic modeling. The block diagram in Figure 1 shows the methodology in brief.

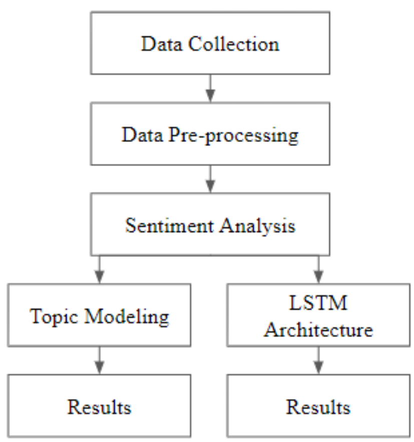

Fig 1. Block Diagram of LSTM and Topic Modeling Methodology

2.1 Data Collection

Customer review data is generated in millions by Airbnb's[17]. This review data can be used to gather insights and understand the nature of reviews. This review data is available on InsideAirbnb.com. For this study, we gather review data for Portland, Sydney, and Singapore from InsideAirbnb.Com. To create one combined dataset for training, the three datasets are concatenated to form a single dataset, which is later split in training and testing sets for modeling purposes.

2.2 Data Preparation and pre-processing

Text cleaning and pre-processing is a necessary step before moving to modeling. It is important to remove unnecessary or non-relevant pieces of information from the data which may affect the modeling and ultimately the performance metrics. Using Natural Language Processing, the reviews were pre-processed and cleaned by performing the following series of steps:

1. Removing stop words
2. Removing hashtags
3. Removing URL's and HTML tags
4. Removing emojis
5. Convert the texts to lower case.

The pre-processing was done by making use of regular expressions. Stop words were removed with the help of nltk stopwords library. The returned processed texts were lemmatized using WordNetLemmatizer. Lemmatization is where a word is converted to its root form. While converting the word, lemmatization is known to the consider the context of the word and then convert it to its root form which is meaningful.

2.3 Sentiment Analysis

scores
{'neg': 1.0, 'neu': 0.0, 'pos': 0.0, 'compound...
{'neg': 0.0, 'neu': 1.0, 'pos': 0.0, 'compound...
{'neg': 0.0, 'neu': 0.633, 'pos': 0.367, 'comp...
{'neg': 0.0, 'neu': 0.0, 'pos': 0.0, 'compound...

Fig 2. Output returned in the form of dictionary

Sentiment analysis on the pre-processed dataset was done using VADER lexicon [18] portrays neutral, positive, or negative emotion/polarityThedataframe with the processed review texts are passed to VADER to get the sentiment scores for each review. The result returnedwas in the form of a dictionary (Figure 2) was split into their columns respectively (Figure 3), for further processing.

neg	neu	pos	compound
1.000	0.000	0.000	-0.5423
0.000	1.000	0.000	0.0000
0.000	0.633	0.367	0.4404

Fig 3. Output split into columns

Data Pre-processing, cleaning and the Sentiment analysis are the common steps taken for the datasets before proceeding to the two parts of research. Sub-section 2.4.A contains implementation details of LSTM and Sub-section 2.4.B contains Topic modeling implementation.

2.3.1 Deep Learning using LSTM

LSTM is a well-known RNN Architecture. LSTM is known to handle the problem of vanishing gradient. A LSTM Unit has three gates. Each unit consists of Forget Gate, the Input Gate, and the Output Gate, respectively. The information flow is controlled by these three gates. To put in simple words, Forget Gate is responsible to decide if the information which has been received should be forgotten or it should be retained. Input gate then learns new pieces of information and output gate sends this updated information to the next LSTM unit. LSTM also has the hidden state and the cell state. The hidden state is called as short-term memory and the cell state is called as long term memory.

After getting the sentiment scores by using VADER, the resulting dictionary was split into respective columns. For the modeling we only considered positive and negative reviews. In this research we create the label column with the help of comparison between VADER returned scores instead of manually labelling each text review as positive or negative. A label column was created, for every row containing review and the sentiment scores for the review. If the positive sentiment is greater than negative sentiment, label is marked 1, and if negative sentiment is greater than or equal to positive sentiment the label is marked 0. The label counts are shown in Table 1. This label column is the output column, on which the LSTM Model was trained.

Table 1. Label Counts

Label	Counts
1 - positive	603885
0 - negative	26332

The train test split was performed, post which the text data was tokenized using Tokenizer from keras library. The sequences were padded with 0 to make them of same length. The maximum length to which the sequences were padded was 300. The vocabulary size obtained was 139380.

As observed from Table 1, the dataset is an imbalanced dataset, the total counts of labels 1 and 0 are unequal and have a concerning difference. Minority examples may get ignored or misclassified compared to the majority examples in machine learning algorithms, due to huge difference between the number of observations[19]. To make the dataset balanced, RandomUnderSampler() from imblearn library was used for under sampling the data to have equal counts for the labels. The counts after resampling the data are given in Table 2.

Table 2. Label counts after Under sampling

Label	Counts
1 - positive	21066
0 - negative	21066

After the processing was done, the LSTM model is trained on the training data. Sequential LSTM model was used for training with 100 units of LSTM, embedding layer, 0.5 dropout, and the resulting dense layer had one unit which used sigmoid activation function. Adam optimizer was used during model compilation. The training data was fit with batch size of 150 for five epochs. The hyperparameter values are shown in Table 3. The summary of model is shown in Figure 4.

Table 3. LSTM Hyperparameters

Parameter	Value	Parameter	Value
Input Length	300	Embedding size	300
Dropout	0.5	LSTM Size	100
Recurrent dropout	0.2	Epochs	5
Batch Size	150	Loss	Binary Cross entropy
Optimizer	Adam	Output Activation	Sigmoid

2.3.2 Topic Modeling

Airbnb hosts should know the specific and influencing factors which are crucial for building and gaining the customer's trust. Understanding the customer satisfaction and dissatisfaction factors is a major step in helping hosts make changes in their respective Airbnb and taking crucial decisions which will help them satisfy customer needs, in return gaining customers trust and satisfaction.

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 embedding (Embedding)       (None, 300, 300)          41814000

 dropout (Dropout)           (None, 300, 300)          0

 lstm (LSTM)                 (None, 100)               160400

 dense (Dense)               (None, 1)                 101

=================================================================
Total params: 41,974,501
Trainable params: 160,501
Non-trainable params: 41,814,000
_________________________________________________________________
```

Fig 4. Model Summary

Topic Modeling was done only on the Singapore review dataset. After completing pre-processing of the texts and applying sentiment analysis on every review, for every listing, the sum of positive and negative sentiment is calculated, and accordingly split into two different dataframe. For all reviews in a listing, when collective positive sentiment sum is greater than collective negative sentiment, the listing review details are added to positive sentiment dataframe. Similar procedure is done to create negative sentiment dataframe. We do this primarily, to have topics for each sentiment, respectively.

The text reviews are vectorized using word2vec word embedding. Topic Modeling was done on both data frames using LDA from genism library. LDA is a probabilistic algorithm introduced by authors[(20)], which is used to discover the hidden topics in text data.

3 Results and Discussion

3.1 Deep Learning using LSTM

The training accuracy was obtained to be of 96.37%. And the testing accuracy was obtained of 93.89% on the testing data. The performance metrics that are considered for the research are recall and F1 score. The higher score of recall can be interpreted as, the model being able to identify higher number of true positives. The higher the f1-score, means the model performance is better.

The performance metrics obtained on the testing dataset are specified in Table 4. 94% and 99% recall score were obtained for positive and negative reviews, respectively. This can be interpreted as; the model is able to predict positive and negative sentiment properly. 97 % and 57% f1-score were obtained for positive and the negative sentiment, respectively. The performance metrics proves that the model is able to differentiate between positive and negative reviews, which were marked based on the scores obtained from sentiment analysis using VADER and not manually.

F1-Score of 81% and 32% was obtained for positive and negative reviews respectively by authors of[(1)] for sentiment analysis using LSTM on a Kaggle dataset. 95% accuracy was obtained by authors[(18)] for binary sentiment classification using LSTM, which used GloVe and word2vec word embeddings, on Yelp dataset. 82% accuracy was obtained for sentiment analysis using RNN on Sentiment140 dataset by authors[(19)]. This study shows improved training and testing accuracy, performance metrics like F1-score and Recall, compared to results from previous studies.

Table 4. Performance Metrics obtained on testing dataset

	Recall	**F1-Score**
1 - positive	0.94	0.97
0 - negative	0.99	0.57

3.2 Topic Modeling

The positive topics that are observed the most were, price, location, cleanliness, and the host. 'Worth price,' 'Value for money,' 'affordable,' were the keywords identified for the 'price' topic. 'Beautiful,' 'view,' sky' were some of the keywords stating the topic 'location.' And, 'responsive,' 'friendly,' 'service' were some keywords specified for the host and the staff of the Airbnb. Regarding the negative topics the most stood out topic was the 'location' of the Airbnb, stating keywords like 'noisy,' 'loud,' 'screaming' etc.

Research results by authors[13] discovered that the host communication was found pleasant by the customers. The findings by authors of[12] and[21], share a common factor for customer satisfaction being the neighborhood the listing is in, which also align with the 'location', topic found in the second part of this paper. Guests are bothered by the loud noise in the surrounding or by the host. The guests like it when the location is a beautiful place. Location and host communication, accommodations were common customer satisfaction and dissatisfaction factors found in the previous studies as mentioned in the introduction section. These factors seem to be commonly occurring for most cities review data. The results from previous studies, align with the results obtained in this paper, which help us to understand the factors affecting satisfaction and also the dissatisfaction of the customers. Host being friendly towards the guest and having good communication with them, makes the guests to feel positive emotions. This study adds more proof to the existing literature, on the factors affecting customer reviews, thus adding value to it. This study also helps to justify that location, and host are major factors, and thus adding more generalizability to the Airbnb review literature.

The studies mentioned in the introduction section suggest that the host should be more aware of the guests and have excellent communication with the guests, because host friendliness, service and staff behavior are the important factors which contribute towards the positive sentiment of the guests and their satisfaction. The results obtained in this paper, align with the results conducted by most studies which state the main factors affecting customer satisfaction and dissatisfaction are, host interaction, the accommodation environment and the amenities provided. The customer satisfaction factors which were discovered in this paper, price, location, host behavior, are found to be major influencing factors for the customer sentiment. The results help us understand that the customer factors like 'price', the 'location' and what it has to offer, the host and staff services are always important factors in whichever city they stay.

4 Conclusion

The results of customer satisfaction and dissatisfaction factors can be helpful to generalize the needs and improvements, for the Airbnb hosts and help improve the customer experience in the future. The hosts should pay specific attention to these commonly occurring factors which are important for customer sentiment. The important factors like price of the Airbnb, location, and the host behavior, found in this study, lie in line with the previous studies conducted by various researchers. The results obtained in this study can be considered as a steppingstone to generalize the customer satisfaction and dissatisfaction factors of sharing economy to a certain extent.

In this paper we trained LSTM model with word embedding vectorizer. The training accuracy after 5 epochs and batch size of 150 was 96.37%, and the testing accuracy was 93.89%. Considering we trained the model for three different cities, and we had a vocabulary of 139380, the performance metrics like F1-score and the recall show good results. This study used a unique approach of converting the data into supervised learning for LSTM modeling, by creating the conditional sentiment label column using VADER returned scores. When compared to previous studies, improved accuracy and performance metrics are obtained in this study. The limitation of the study was the computing resources limitation. With stronger computing resources and by combining variety of datasets, which can create huge vocabulary for training, can help improve results. Future work can target, training deep learning models by combining variety of review datasets and building vocabulary from various sources, to make the model stronger, which might help to understand factors influencing the change in performance metrics.

References

1) Raza MR, Hussain W, Varol A. Performance Analysis of Deep Approaches on Airbnb Sentiment Reviews. *2022 10th International Symposium on Digital Forensics and Security (ISDFS)*. 2022. Available from: https://doi.org/10.1109/ISDFS55398.2022.9800816.
2) Alsurayyi WI, Alghamdi NS, Abraham A. Deep Learning with Word Embedding Modeling for a Sentiment Analysis of Online Reviews. *International Journal of Computer Information Systems and Industrial Management Applications*. 2019;11:227–241. Available from: http://www.mirlabs.org/ijcisim/regular_papers_2019/IJCISIM_22.pdf.
3) Dang NC, Moreno-García MN, Prieta FDL. Sentiment Analysis Based on Deep Learning: A Comparative Study. *Electronics*. 2020;9(3):483. Available from: https://doi.org/10.3390/electronics9030483.
4) Sutherland I, Kiatkawsin K. Determinants of Guest Experience in Airbnb: A Topic Modeling Approach Using LDA. *Sustainability*. 2020;12(8):3402. Available from: https://doi.org/10.3390/su12083402.
5) Lima JIMP, Pessanha GRG, Araujo MVP, Alves RC, De A, Ces MFP, et al. Place branding Pernambuco: analysis of the feelings of the users through Instagram hashtags". *Brazilian Journal of Marketing*. 2022;21(1):154–184. Available from: https://doi.org/10.5585/remark.v21i1.20578.
6) Chiny M, Bencharef O, Hadi MY, Chihab Y. A Client-Centric Evaluation System to Evaluate Guest's Satisfaction on Airbnb Using Machine Learning and NLP. *Applied Computational Intelligence and Soft Computing*. 2021;2021:1–14. Available from: https://doi.org/10.1155/2021/6675790.
7) Chiny M, Bencharef O, Chihab Y. Towards a Machine Learning and Datamining approach to identify customer satisfaction factors on Airbnb. *2021 7th International Conference on Optimization and Applications (ICOA)*. 2021;p. 1–5. Available from: https://doi.org/10.1109/ICOA51614.2021.9442657.

8) Keawtoomla N, Pongwat A, Bootkrajang J. Using Latent Dirichlet Allocation to investigate guest experience in Airbnb accommodation during COVID-19 pandemic in the United Kingdom. *2022 19th International Joint Conference on Computer Science and Software Engineering (JCSSE)*. 2022;p. 1–6. Available from: https://doi.org/10.1109/JCSSE54890.2022.9836314.
9) Piris Y, Gay AC. Customer satisfaction and natural language processing. *Journal of Business Research*. 2021;124:264–271. Available from: https://doi.org/10.1016/j.jbusres.2020.11.065.
10) Li Z, Chen H, Huang X. Airbnb or Hotel?: A Comparative Study on the Sentiment of Airbnb Guests in Sydney - Text Analysis Based on Big Data. *International Journal of Tourism and Hospitality Management in the Digital Age (IJTHMDA)*. 2020;4(2):1–10. Available from: https://doi.org/10.4018/IJTHMDA.2020070101.
11) Santos AIGP, Perinotto ARC, Soares JRR, Mondo TS, Cembranel P. Expressing the Experience: An Analysis of Airbnb Customer Sentiments. *Tourism and Hospitality*. 2022;3(3):685–705. Available from: https://doi.org/10.3390/tourhosp3030042.
12) Ding K, Choo WC, Ng KY, Ng SI, Song P. Exploring Sources of Satisfaction and Dissatisfaction in Airbnb Accommodation Using Unsupervised and Supervised Topic Modeling. *Frontiers in Psychology*. 2021;12:659481. Available from: https://doi.org/10.3389/fpsyg.2021.659481.
13) Vassilikopoulou A, Kamenidou I, Priporas CV. Negative Airbnb reviews: an aspect-based sentiment analysis approach. *EuroMed Journal of Business*. 2022. Available from: https://doi.org/10.1108/EMJB-03-2022-0052.
14) Madhi HAB, Alhammah MM. What Drives Airbnb Customers' Satisfaction in Amsterdam? A Sentiment Analysis. *International Journal of Advanced Computer Science and Applications*. 2021;12(6). Available from: https://doi.org/10.14569/IJACSA.2021.0120628.
15) Zhang G, Cui R, Cheng M, Zhang Q, Li Z. A comparison of key attributes between peer-to-peer accommodations and hotels using online reviews. *Current Issues in Tourism*. 2020;23(5):530–537. Available from: https://doi.org/10.1080/13683500.2019.1575339.
16) Marine-Roig E. Analytics in hospitality and tourism: Online travel reviews. *Advances in Hospitality and Tourism Information Technology*. 2021;p. 1–27. Available from: https://www.researchgate.net/publication/358817124_Analytics_in_hospitality_and_tourism_Online_travel_reviews.
17) Joseph G, Varghese V. Analyzing Airbnb Customer Experience Feedback Using Text Mining. In: Sigala, M, Rahimi, R, Thelwall, M, editors. Big Data and Innovation in Tourism, Travel, and Hospitality. Springer Singapore. 2019;p. 147–162. Available from: https://doi.org/10.1007/978-981-13-6339-9_10.
18) Hutto C, Gilbert E. VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text. *Proceedings of the International AAAI Conference on Web and Social Media*. 2014;8(1):216–225. Available from: https://doi.org/10.1609/icwsm.v8i1.14550.
19) Hamad RA, Kimura M, Lundström J. Efficacy of Imbalanced Data Handling Methods on Deep Learning for Smart Homes Environments. *SN Computer Science*. 2020;1(4):204. Available from: https://doi.org/10.1007/s42979-020-00211-1.
20) Blei DM, Ng AY, Jordan MI. Latent dirichlet allocation. *Journal of machine Learning research*. 2003;3:993–1022. Available from: https://www.jmlr.org/papers/volume3/blei03a/blei03a.pdf.
21) Daniel S, Saidat A, Sanni L, Rhue. A Power-threat View of The Role of Neighborhood Demographics on Airbnb Review Sentiments. *AMCIS 2022 Proceedings*. 2022. Available from: https://aisel.aisnet.org/amcis2022/sig_si/sig_si/1.