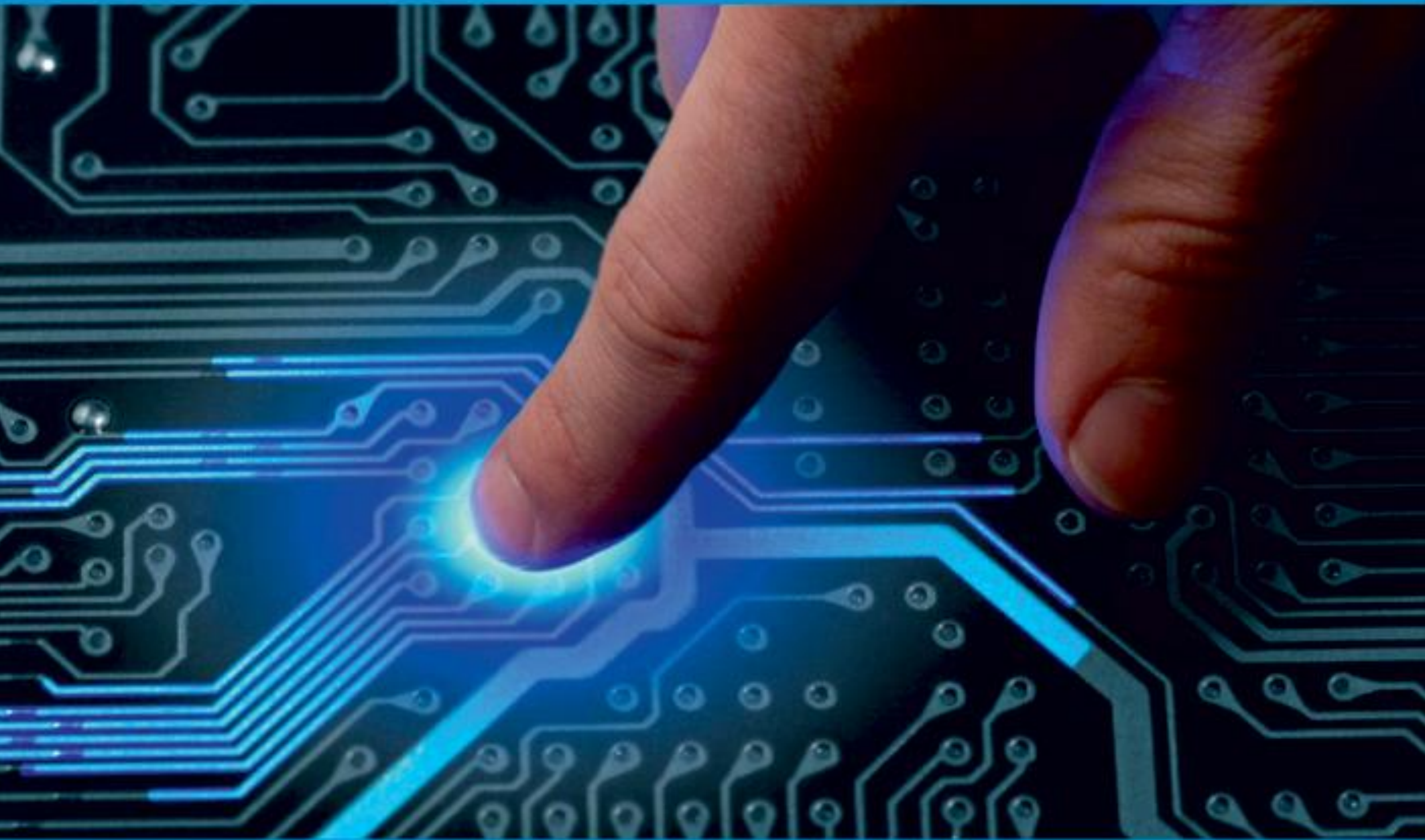
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Special Issue 1, March 2024

1st International Conference on Machine Learning, Optimization and Data Science

Organized by

Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur, India

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 8.379

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

An Efficient Way for Detecting Bad Customer Reviews Using NLP

Vishal Paranjape, Saurabh Sharma

Dept. of Computer Science & Engineering, Baderia Global Institute of Engineering & Management, Jabalpur, India

ABSTRACT: In today's world of digitization and high availability of internet reviews given by customer play a very important role in defining the next buying pattern of a customer. A number of companies like amazon, flipkart etc provides platform for their customers to give their real experience about the product so that they can boost up the sales and be able to identify their prospective customers. Natural Language Processing plays an important role in identifying customers based on their positive or negative review. In this paper we are deploying NLP techniques for identifying the positive or negative sentiment using sentiment analysis.

KEYWORDS: sentiment analysis, natural language processing,

I. INTRODUCTION

The amount of text data generated has significantly increased during the past ten years along with the tremendous growth in internet usage. Analyzing text data is crucial as it is available today in articles, reviews, and online discussions on several topics[1]. One of the key uses of text data analysis is to ascertain how customers feel about a particular service.

Natural language processing (NLP) uses a technique called sentiment analysis that includes extracting emotions from some raw texts. Sentiment analysis is used to analyze post on social media to see for positive or negative sentiments of user [2]. In other words, sentiment analysis is the classification of opinions about a specific product, service, or topic in a text (word, sentence, or document) in order to computationally ascertain the writer's polarity or attitude toward the topic — positive, negative, or neutral. (Fatmeh Hemmatian and Muhammad Karim Sohrabi) (Mohammad Karim Sohrabi and Fatmeh Hemmatian). Two major things that play a crucial role for creating sentiment analysis model is data preparation and feature extraction as model accuracy is obtained by feature extraction. Data preprocessing methods include noise reduction, normalization, tokenization, and vectorization, among others. (Vikram Singh and Balwinder Saini, 2014).

The proposed effort will investigate the effects of various preprocessing and feature extraction procedures on the dataset of restaurant reviews in order to understand how they impact the model's accuracy [3]. The necessary results have been produced using text analysis, natural language processing, automatic classification approaches, and text classification methods.

The paper is structured as follows: Section 2 covers the related work that has been done. In Section 3 of this article, methodology has been discussed. Section 4 presents experimental work as well as the analysis report. Section 5 brings the assignment to an end with conclusion.

II. RELATED WORK

A brief survey about the related work performed on Opinion Mining in yesteryears has been explained in this section. There have been many advancements and researches carried in the field of opinion mining and Sentiment Analysis in the past years [4].

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

Table 1: Comparison between various researchers and their work

Authors	Title	Work on	Algorithms Used	Accuracy Obtained
Elmogy, Ahmed & Tariq, Usman & Mohammed, Ammar & Ibrahim, Atef. [5]	Fake Reviews Detection using Supervised Machine Learning. International Journal of Advanced Computer Science and Applications	Worked on Textual & behavioural features of review.	Support Vector Machine (SVM), Decision Tree, and K-Nearest Neighbors	88%
R. Barbado, O. Araque, and C. A. Iglesias [6]	A framework for fake review detection in online consumer electronics retailers	generate the dataset from Yelp through Yelp scrapping	fake feature framework for extraction and characterization used	82%
E. I. Elmurngi and A.Gherbi [7]	Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques	WEKA tool was used for implementing	SVM, Decision Trees, Logistic Regression, and Naïve Bayes	81.61%
Monica, C., Nagarathna [8]	Detection of Fake Tweets Using Sentiment Analysis	sentiment analysis, they gave the sentiment score based on the lexicon features	Multi-layer perceptron (MLP), Decision Trees, and Random forest algorithms	81%
Mohawesh, Rami & Xu, Shuxiang & Tran, Son & Ollington [9]	Fake Reviews Detection: A Survey	Parts of Speech (PoS), Bag of Word (BoW), Linguistic inquire and word count,	Human Methods, the Amazon Mechanical Turk method, and RULR based method	70.2%

III. METHODOLOGY

The information from hotel reviews are used in our work to check for the sentiments of customers on the basis of their reviews in terms of text. The reviews in terms of text message consist of the experience any customer have in hotel along with textual description and rating. We are using the following hotel dataset available via the kaggle link given below: https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-Europe.

On the basis of liking of the customer they giv reviews that are either positive which means customer is happy with the services or negative which shows the dissatisfaction of the customer. Scores of reviews can be between 2.5 and 10.

To make the issue simpler, we'll divide those into two groups:

- The average rating for negative reviews is 5.
- Reviews with high overall rankings (>= 5).

This project aims to demonstrate the use of Python for sentiment analysis. Some of the key libraries we'll use are thefollowing:

- NLTK, the most well-known Python NLP module
- Gensim: a toolbox for subject modeling and vector space modeling
- An important library for Python in ML is Scikit-learn.

Every data in the given dataset comprise of single guest evaluation for a single hotel. To make the issue simpler, we'lldivide those into two groups:

- If overall rating < 5 means bad reviews
- good reviews have overall ratings >= 5

- The average rating for negative reviews is 5.
- Reviews with high overall rankings (>= 5).

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

	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

Initial dataset

Figure 1: Analysis of Good Review & Bad Review from Initial Dataset

We use our unique 'clean_text' function, which carries out the following modifications, to clean up textual data:Text should be lowered, tokenized (divided into words), and punctuation removed.

- exclude pointless words that include numerals.
- removal of stop words like "is,"a","the", "this," etc.
- To tag each word according to its Part-Of-Speech (POS) categorization, use the WordNet lexical database.
- Lemmatize the text by changing each word to its corresponding root (e.g., rooms to room, slept to sleep).

Feature engineering

Mostly customer reviews are based on the eminities and luxuries that were provided to them during their stay at a particular hotel and how they felt on their stay. Vader, a sentiment analysis feature of NLTK was applied in our work. Vader through his library determined the positive and negative words. Sentence context is also used for sentiment scores.

4 values are returned by Vader:

- a neutrality score
- a +ve score
- a -ve score
- an overall score

For every word and document we use the TF-IDF (Term Frequency — Inverse Document Frequency). But why not just keep track of how often each term appears in each document? This method's drawback is that it ignores the relative importance of each word in the texts.
It is doubtful that a word that appears in almost every text can offer useful information for analysis. On the other hand, uncommon terms could be far more ambiguous.

This issue is resolved by the TF-IDF measure, which computes the relative value of a word based on how many texts it can be discovered in addition to the traditional number of times it appears in the text.

We add TF-IDF columns for each word for the purpose to limit the amount of output and filter out those words that

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

appear in at least ten different texts,

Exploratory data analysis

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

Figure 2: Score of Good Review & Bad Review

Because less than 5% of our evaluations are classified as unfavorable, our dataset is severely skewed. The modeling portion will benefit greatly from this knowledge.

Figure 3: WordCloud from the customer reviews

IV. EXPERIMENTAL RESULT

The majority of the words—room, staff, breakfast, etc.—are in fact associated with hotels. Perfect, adored, pricey, detest, and other terms like these are more closely associated with how guests felt throughout their hotel stay.

	review	pos
43101	A perfect location comfortable great value	0.931
211742	Clean comfortable lovely staff	0.907
175551	Friendly welcome Comfortable room	0.905
365085	Good location great value	0.904
109564	Clean friendly and comfortable	0.902
145743	Good value amazing location	0.901
407590	breakfast excellent Clean comfort	0.899
407546	Great place I enjoyed	0.881
218571	Beautiful Quirky Comfortable	0.878
436901	Lovely comfortable rooms	0.877

Highest positive sentiment reviews

Figure 4: Highest positive sentiment reviews

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

Good feedbacks are associated with most positive reviews.

	review	neg
193086	No dislikes LOCATION	0.831
318516	A disaster Nothing	0.804
29666	A bit noisy No	0.796
426057	Dirty hotel Smells bad	0.762
263187	Very bad service No	0.758
181508	Window blind was broken	0.744
174178	no bad experience location	0.740
291281	nothing great clean comfortable quite hotel	0.733
233344	It was awful No	0.722
384771	Very bad atmosphear noisy weird smells unfrie...	0.713

Highest negative sentiment reviews

Figure 5: Highest negative sentiment reviews

Several mistakes can be identified in the most unfavorable reviews: Vader frequently interprets "no" and "nothing" as negative statements.

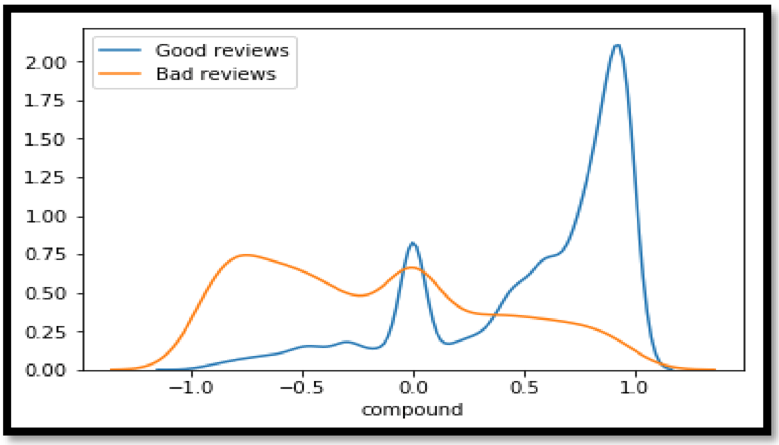


Figure 6: Sentiment distribution

The distribution of positive and negative review sentiment is depicted in the graph up top. We can see that Vader regardsmost of them as having received highly positive assessments.

This demonstrates to us the significance of previously computed sentiment features for our modeling section. The features we wish to employ to train our model are first selected. Next, we divided our data into two groups:

- One to test the performance of our model; and

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 8.379 || Monthly Peer Reviewed & Refereed Journal |

| "1st International Conference on Machine Learning, Optimization and Data Science"|

Organized by

| Department of Computer Science and Engineering, Baderia Global Institute of Engineering and Management, Jabalpur|

|| Volume 12, Special Issue 1, March 2024 ||

Next, we will make predictions using a Random Forest (RF) classifier.

	feature	importance
3	compound	0.037007
2	pos	0.025501
0	neg	0.024003
10	doc2vec_vector_4	0.020557
6	doc2vec_vector_0	0.018240
8	doc2vec_vector_2	0.017705
9	doc2vec_vector_3	0.017626
7	doc2vec_vector_1	0.016926
4	nb_chars	0.016663
1	neu	0.014799
5	nb_words	0.014324
950	word_dirty	0.010294
2239	word_nothing	0.009593
2853	word_room	0.009208
285	word_bad	0.008143
3202	word_staff	0.006811
1639	word_hotel	0.006578
1945	word_location	0.006349
3216	word_star	0.006328
2284	word_old	0.006206

Most important features

Figure 7: Most important features

The elements that originate from the previous sentiment analysis are in fact the most crucial ones.

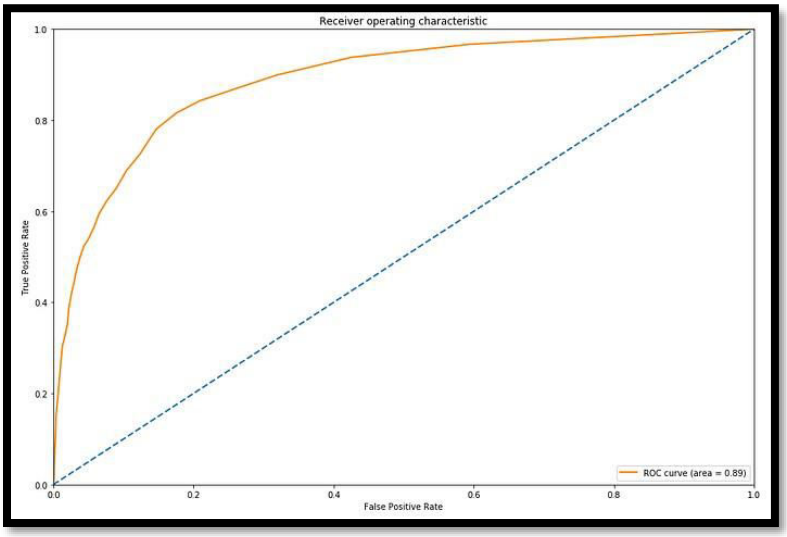


Figure 8: ROC Curve

A useful graph to summarize the effectiveness of our classifier is the ROC (Receiver Operating Characteristic) curve. The accuracy of the forecasts increases as the curve rises over the diagonal baseline.

Because of the imbalance in our dataset, the amount of negatives in this case correlates to the high number of positive evaluations. This means that our FPR will typically remain extremely low, even with some False Positives. Our model will be able to predict many false positives while still having a low false positive rate.

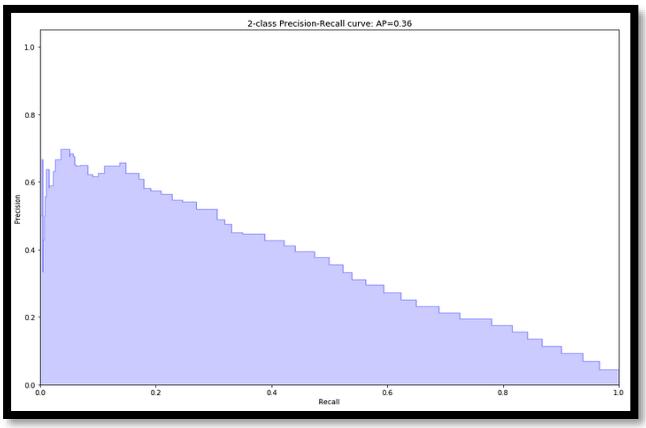


Figure 9: PR Curve

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

We can observe that as recall is raised, precision falls. This indicates that we need to choose a prediction threshold that satisfies our needs. In order to get a high recall, we should select a low prediction threshold that, even with reduced accuracy, will allow us to find most observations in the positive class. Instead, we should select a high threshold that will give us a high precision and a low recall if we want to be highly sure in our predictions but don't care if we don't locate all the positive observations.

A classifier with this level of precision would have a positive observation rate of 4.3%. For any recall value, the precision would stay the same, providing an AP of 0.043. The apparent power (AP) of our model is around 0.35, more than eight times higher than the apparent power of the random technique. This demonstrates the strong predictive capability of our model.

V. CONCLUSION

For making predictions we use raw text as an input in our work. Extracting the important features from the raw text is the most important and challenging task. For the purpose of extracting more learning features we can use this data as a good source to increase the predictive power of models as well.

REFERENCES

[1] C. Silva and B. Ribeiro, "The importance of stop word removal on recall values in text categorization," in Neural Networks, 2003. Proceedings of the International Joint Conference on, vol. 3. IEEE, 2003, pp. 1661–1666.

[2] J. Plisson, N. Lavrac, D. Mladeni´c et al., "A rule based approach to word lemmatization," 2004.

[3] Baishya, D., Deka, J.J., Dey, G. et al. SAFER: Sentiment Analysis Based FakE Review Detection in E-Commerce Using Deep Learning. SN COMPUT. SCI. 2, 479 (2021).

[4] G. Fei, A. Mukherjee, B. Liu, M. Hsu, M. Castellanos, and R. Ghosh, "Exploiting burstiness in reviews for review spammer detection," in Seventh international AAAI conference on weblogs and social media, 2013.

[5] Elmogy, Ahmed & Tariq, Usman & Mohammed, Ammar & Ibrahim, Atef. (2021). Fake Reviews Detection using Supervised Machine Learning. International Journal of Advanced Computer Science and Applications. 12.10.14569/IJACSA.2021.0120169.

[6] R. Barbado, O. Araque, and C. A. Iglesias, "A framework for fake review detection in online consumer electronics retailers," Information Processing & Management, vol. 56, no. 4, pp. 1234 – 1244, 2019.

[7] E. I. Elmurngi and A.Gherbi, "Unfair Reviews Detection on Amazon Reviews using Sentiment Analysis with Supervised Learning Techniques," Journal of Computer Science, vol. 14, no. 5, pp. 714–726, June 2018.

[8] Monica, C., Nagarathna, N. Detection of Fake Tweets Using Sentiment Analysis. SN COMPUT. SCI. 1, 89 (2020).

[9] Mohawesh, Rami & Xu, Shuxiang & Tran, Son & Ollington, Robert & Springer, Matthew & Jararweh, Yaser & Maqsood, Sumbal. (2021). Fake Reviews Detection: A Survey. IEEE Access. 10.1109/ACCESS.2021.3075573.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING