GR-118

Sentiment Analysis On Hotel Reviews

Aravinda P., Dilip Kumar, and Navya P. Advisor: Dr. Mahmut Karakaya

Abstract

- As part of this project, we implemented sentiment analysis.
- It is the use of natural language processing, text analysis, computational linguistics, and biometrics to systematically identify, extract, quantify, and study affective states and subjective information.
- Sentiment analysis is widely applied to voice of the customer materials such as reviews and survey responses, online and social media for applications that range from marketing to customer service to clinical medicine.
- We used the dataset of around 500k records and divided that sample into 8:2 ratio, in training and test dataset, respectively.

Introduction

- In this project we are implementing Sentiment Analysis which is an NLP technique to give sentiment score on text reviews.
- Sentiment analysis works by breaking a message down into topic chunks and then assigning a sentiment score to each topic

Research Question(s)

- What made you choose sentiment analysis for this project?
- How this project useful in real world?

Materials and Methods

- We split the raw data into positive and negative sets and sample the data to speed up the computation.
- We clean the data based on the textual syntax like removing punctuations and numbers.
- We perform sentiment analysis with the help of VADER in NLTK module. VADER evaluates the text of a message and gives you sentiment scores along with positive and negative feedback assessment.

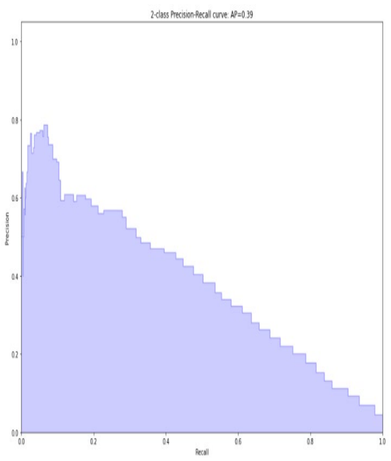
Results

Positive Review and Negative Review Feedback

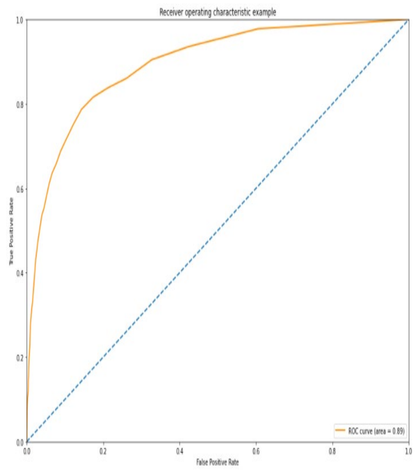
	review	pos
43101	A perfect location comfortable great value	0.931
211742	Clean comfortable lovely staff	0.907
175551	Friendly welcome Comfortable room	0.905
365085	Good location great value	0.904
109564	Clean friendly and comfortable	0.902
145743	Good value amazing location	0.901
407590	breakfast excellent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

	review	neg
193086	No dislikes LOCATION	0.831
356368	Nothing Great helpful wonderful staff	0.812
318516	A disaster Nothing	0.804
458794	Nothing Excellent friendly helpful staff	0.799
29666	A bit noisy No	0.798
426057	Dirty hotel Smells bad	0.762
263187	Very bad service No	0.758
443796	Nothing perfect	0.750
181508	Window blind was broken	0.744
175316	Nothing Super friendly staff	0.743

PR Curve



ROC Curve



Word-Cloud: To have a glimpse at what kind of words appear in our reviews

Conclusions

For each textual review, we predict if it corresponds to a good review (the customer is happy) or to a bad one.
It is totally conceivable to utilize raw text as input for making expectations. The main thing is to have the option to extract the important features from this raw source of information. Such an information can regularly come as a decent integral source in data science projects to extract additional learning elements and increment the predictive power of the models.

Acknowledgments

We would like to express our special thanks of gratitude to Dr. Mahmut Karakaya for guiding us in this project and College of Computing and Software Engineering for giving us this opportunity to work on this project “Sentiment analysis on hotel reviews”, which also helped us in doing a lot research.
Moreover, we would like to thank our ML class and friends who have helped us with their valuable suggestions and guidance and have been very helpful in various stages of project completion.

Contact Information

npallap1@students.kennesaw.edu

apandana@students.kennesaw.edu

dkumar3@students.kennesaw.edu
https://www.linkedin.com/in/dilip-kumar-27069167/

References

[1]Dataset- https://www.kaggle.com/jiashenliu/515k-hotel-reviews-data-in-europe

[2] https://www.ibm.com/cloud/learn/natural-language-processing

[3] https://en.wikipedia.org/wiki/Natural_language_processing

KENNESAW STATE UNIVERSITY
COLLEGE OF COMPUTING AND SOFTWARE ENGINEERING