UNIVERSITY OF
ILLINOIS CHICAGO

Review – Response Analysis using Deep Learning Models

IDS 576 – Deep learning & Applications

AIDA SANATIZADEH, ANIRUDHA BALKRISHNA, SHUBHAM KHODE & VAMSY TAMMINEEDI

Introduction

- Nearly 9 in 10 consumers use online review when purchasing a product.
- Customer reviews could offer tremendous insights into what customers like and dislike
- Responding to online reviews is also important and can help solidify online reputation

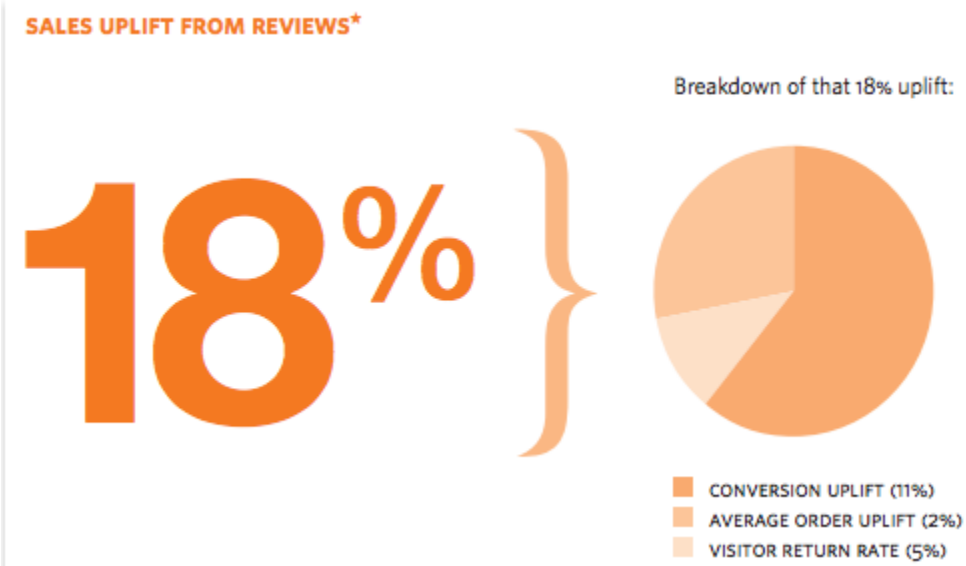


Image Source: https://blog.reevoo.com/ratings-reviews-landscape/

Data

- Use BeautifulSoup to crawl the data from the website: TripAdvisor.com
- Collect 1,467,919 reviews spanning the years 2018, 2021
- All reviews are for 225 chain hotels in the USA
- Out of these reviews, only 123k reviews have manager response (8% of all reviews)

OBJECTIVE

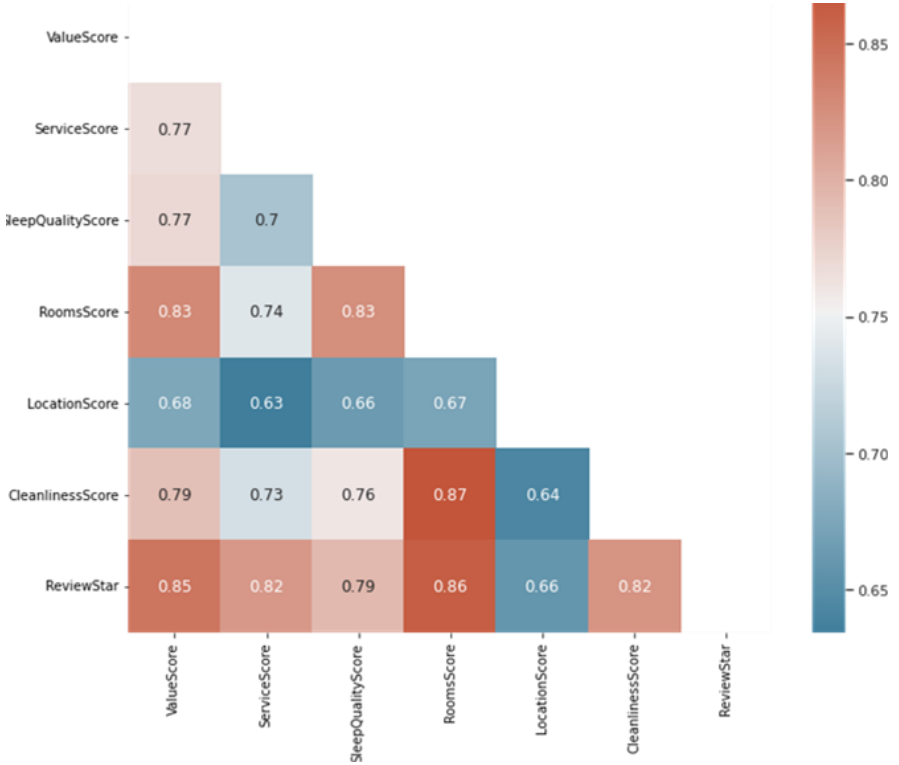
1. Understand the emotions of the reviewers & build a model to determine the criticality of a review with the newly generated emotion as a feature

1. Compare the review and responses to understand whether the reviews were addressed correctly

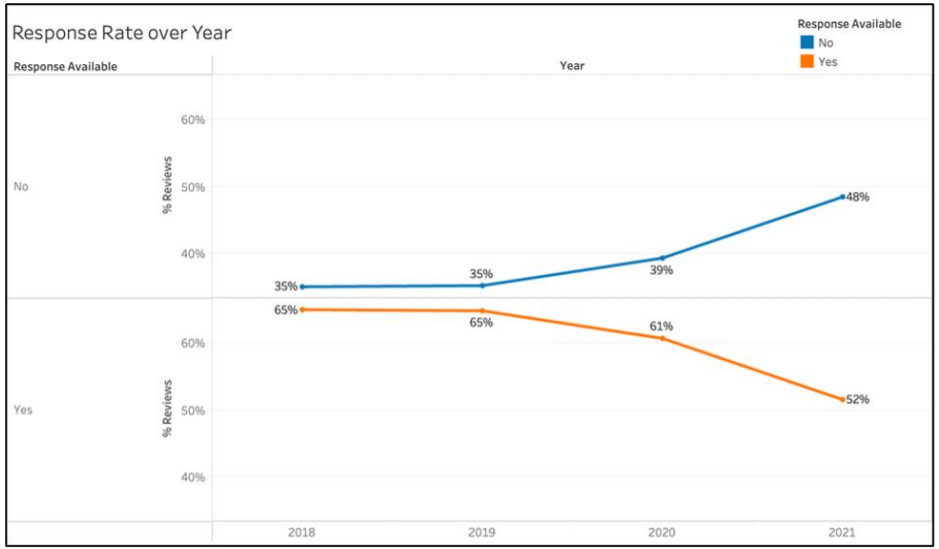
Exploratory Analysis (1)

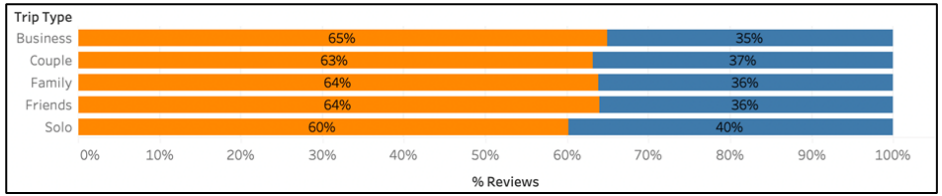
Correlation Analysis between Hotel features and Review Star

- Hotel value score,
- Hotel service,
- Hotel sleep quality score,
- Hotel room score,
- Hotel location score,
- Hotel cleanliness score,

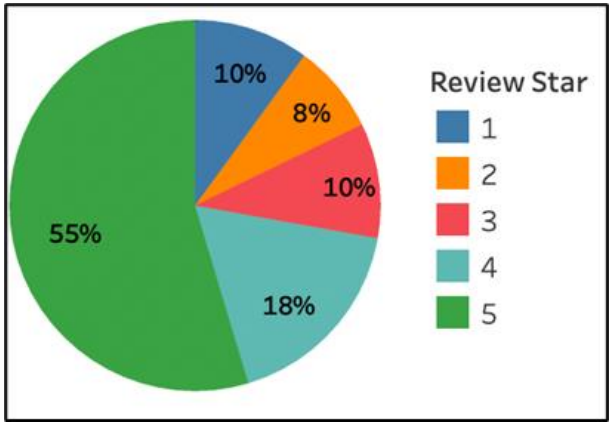


Exploratory Analysis (2)





- Response Rate decreasing over years possible impact of COVID-19
- Business Trip Type - highest response
- Reviews - 4 & 5 star reviews > 70%



OBJECTIVE (1): Criticality Measure For Responses

The review text is used as an input to a pre-trained model trained and fine-tuned using a wide range of emotions dataset to classify 6 different emotions – anger, sadness, fear, surprise, joy and love.

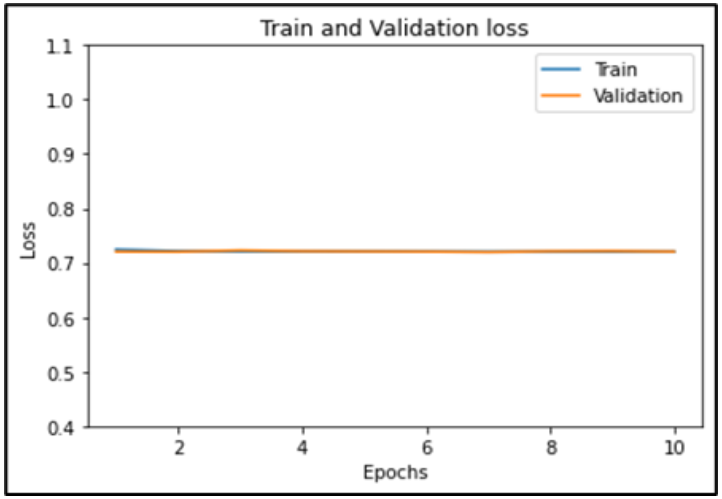
This pre-trained model is fine-tuned on Google T5 Engine which can either output a class label or a span of the input.

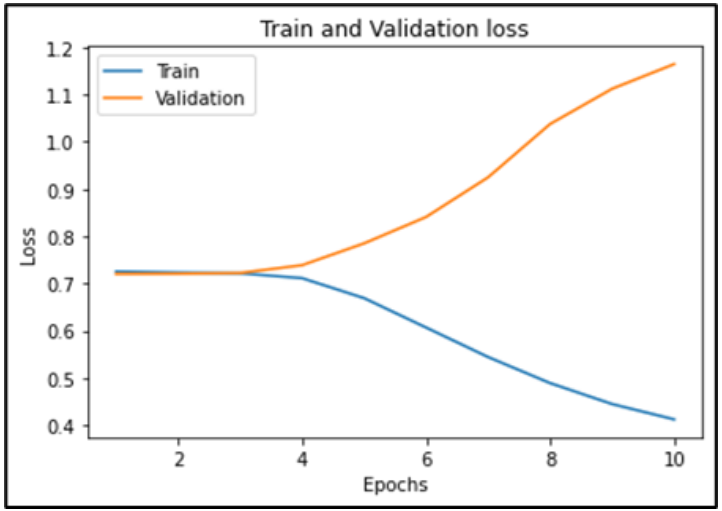
Emotion	Criticality Index
Sadness or Anger	0
Surprise or Fear	1
Joy or Love	2

Emotion	Avg. Response Delay (Days)
Anger	**31.55**
Fear	**29.91**
Joy	**31.04**
Love	**28.18**
Sadness	**29.69**
Surprise	**34.04**

OBJECTIVE (1) RESULTS

- Split into training and validation sets with 30% data as validation data.
- BERT base uncased model is used here to predict the criticality of a given review.
- BERT base has a total of 12 attention heads (encoders) with 768 hidden layers.
- AdamW is used for optimizer
- The model is fine-tuned using the recommended parameters with batch size 16
- Learning rate of – 1e-5(second graph) and – 2e-5 (first graph) are used to fine-tune the model for 10 epochs
- Accuracy : 0.51 and F1 -Score : 0.59





Testing and Interpreting the Results:

One of the reasons for the misclassification could be the label imbalance.

‘Low Criticality’ labels are the majority class

A stratified sampling or scraping more data from TripAdvisor could help us achieve better results.

```
[81] txt = "Nice place to stay. Food was great. I would not hesitate to go back again"
     predict_test(txt)

     'Low Criticality'

[85] txt = 'Not many options are avaialble to eat. I am happy to see that atleast they provide dessert.' # misclassified
     predict_test(txt)

     'Low Criticality'

[92] txt = 'Hotel is not secure!.'
     predict_test(txt)

     'Medium Criticality'

     txt = "I am surprised to see that they don't offer any starters" # misclassified
     predict_test(txt)

     'Low Criticality'

[87] txt = "Food is very bad. Never coming back here again"
     predict_test(txt)

     'High Criticality'
```

OBJECTIVE (2): Similarity Measure

- To understand if the response correctly addresses the review
- We use a BERT (Bidirectional Encoder Representations from Transformers) model to vectorize the texts
- We measure the cosine distance between the two embeddings or vectors
- Higher Similarity indicates relevant context

Testing and Interpreting the Results:

```
sentences[2]
```

Review

```
'It's a great place to stay. Rooms are always super clean . Breakfast is great .
The gym is awesome and u can reserve it by the hour . The grill works great and
and clean'
```

```
sentences[12]
```

Response

```
'Thank you for participating in the TripAdvisor survey. I am happy to hear you e
a positive and memorable guest experience to our guests. I am glad to hear the s
ard to seeing you again! Have a great day and Safe Travels. '
```

```
from sklearn.metrics.pairwise import cosine_similarity
sim = cosine_similarity([sentence_embeddings[2]],[sentence_embeddings[12]])
print((sim*100))
```

```
[[77.010445]]
```

```
[ ] sentences[9]
```

Review

```
'Bad service bad knowledge and cleanliness wasn't the best
log of what was what bad deal want my cash back or 3day cr
ecause I reported it bout how they was not trained right
```

```
[ ] sentences[19]
```

Response

```
'Thank you for the feedback'
```

```
[ ] sim1 = cosine_similarity([sentence_embeddings[9]],[senten
    print(sim1*100)
```

```
[[2.3859882]]
```

Future research

CRM requires learning the managerial responses of high-quality hotels. Responding to a review may increase transaction and labour expenses, while failing to reply to a review may result in lost customer retention possibilities.

Therefore, as an additional objective we can create a bot that will automatically respond to reviews based on the criticality and evaluate its response based on the similarity score. This can be achieved by using Google T5 Engine.

Conclusion

In this project, we have attempted to create a robust method to prioritize reviews and evaluate responses

To prioritize reviews, we extract the emotion associated with the review using deep learning models and then create a criticality index to determine which review is detrimental to the company's image and has to be acted on immediately.

Another part of this project is to evaluate responses. We use deep learning models to embed review and response texts to understand whether a response accurately addresses the review. This will help a company understand if responders are doing a good job of answering reviews.

In summary, criticality helps us determine the priority of responses and similarity helps us determine the correctness of responses.

Thank you