

Aspect-Based Sentiment Analysis

Project Update

Piumi Abeyrathne, Protogene Hahirwabayo, Sifaal Ndandala. April, 9th 2024

Recall: What is Aspect-Based Sentiment Analysis

Sample Hotel Review:

The hotel service was fast and efficient. We got checked in in under 5 minutes and had all our questions answered. However the food was terrible. Breakfast ran from 7-9 am and there was little variety. The wifi was also not very fast and we couldn't stream our favorite shows on TV. Overall the experience was ok.

Traditional Sentiment Analysis

Overall Sentiment:

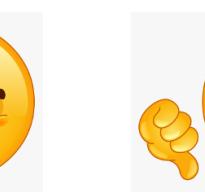


Aspect-Based Sentiment Analysis

Check In Service



Food

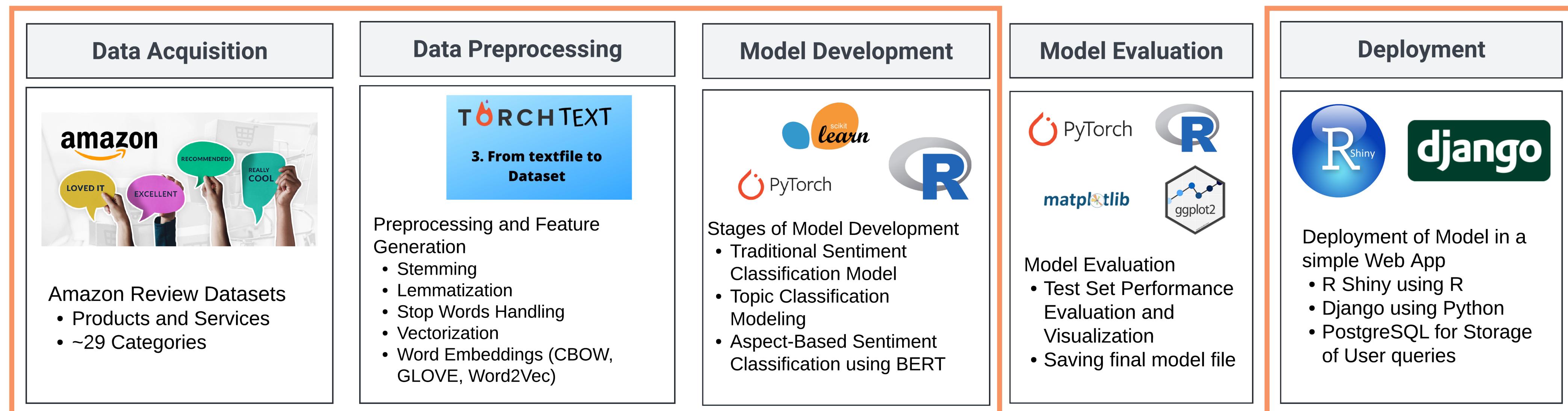


Wifi



Project Update: Where we are?

Aspect-Based Sentiment Classification: Implementation Stages



Project Progress Demo

Text Analyzer	Technical Notes
Enter Your Sample Text:	
the food was truly amazing	
Your Predicted Rating: 4 (Very Positive)	

ABSA Sentiment Analyzer

Url: <https://absa-analyzer-production.up.railway.app/>

How Does it Work?

User Input Text:

Preprocessing:

[punctuation removal, tokenization, and lemmatization]

Feature Generation:

[1-gram model]

Sentiment Model Prediction:

(0-4) [very bad - very good]

The hotel service was fast and efficient. We got checked in in under 5 minutes and had all our questions answered. However the food was terrible. Breakfast ran from 7-9 am and there was little variety. The wifi was also not very fast and we couldn't stream our favorite shows on TV. Overall the experience was ok.

hotel service fast efficient get check minute question answer food terrible breakfast run little variety wifi fast couldn t stream

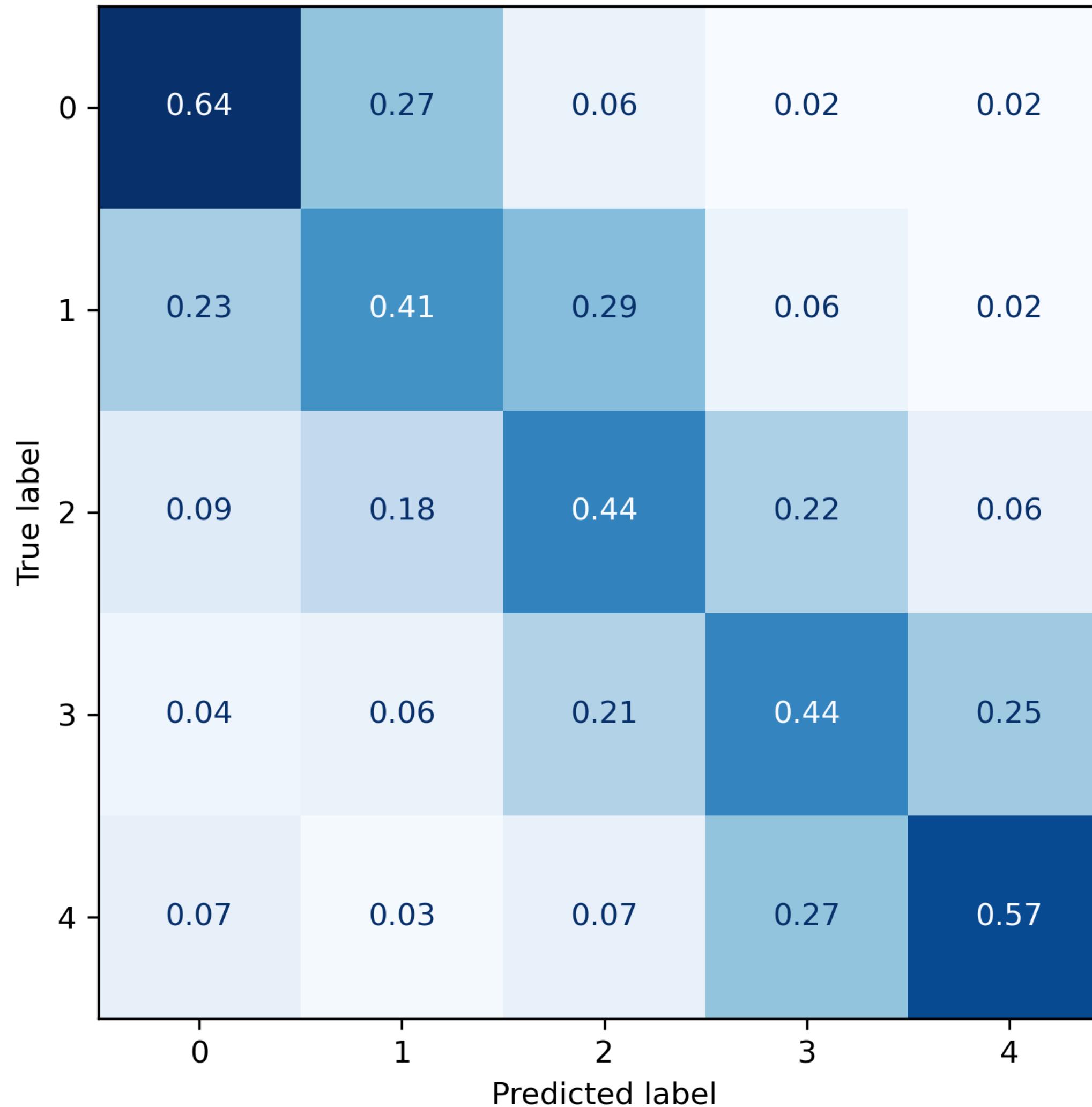
TFIDF Vectorizer: [[0. 0. 0. ... 0. 0. 0.]]

Vector Size: 14102

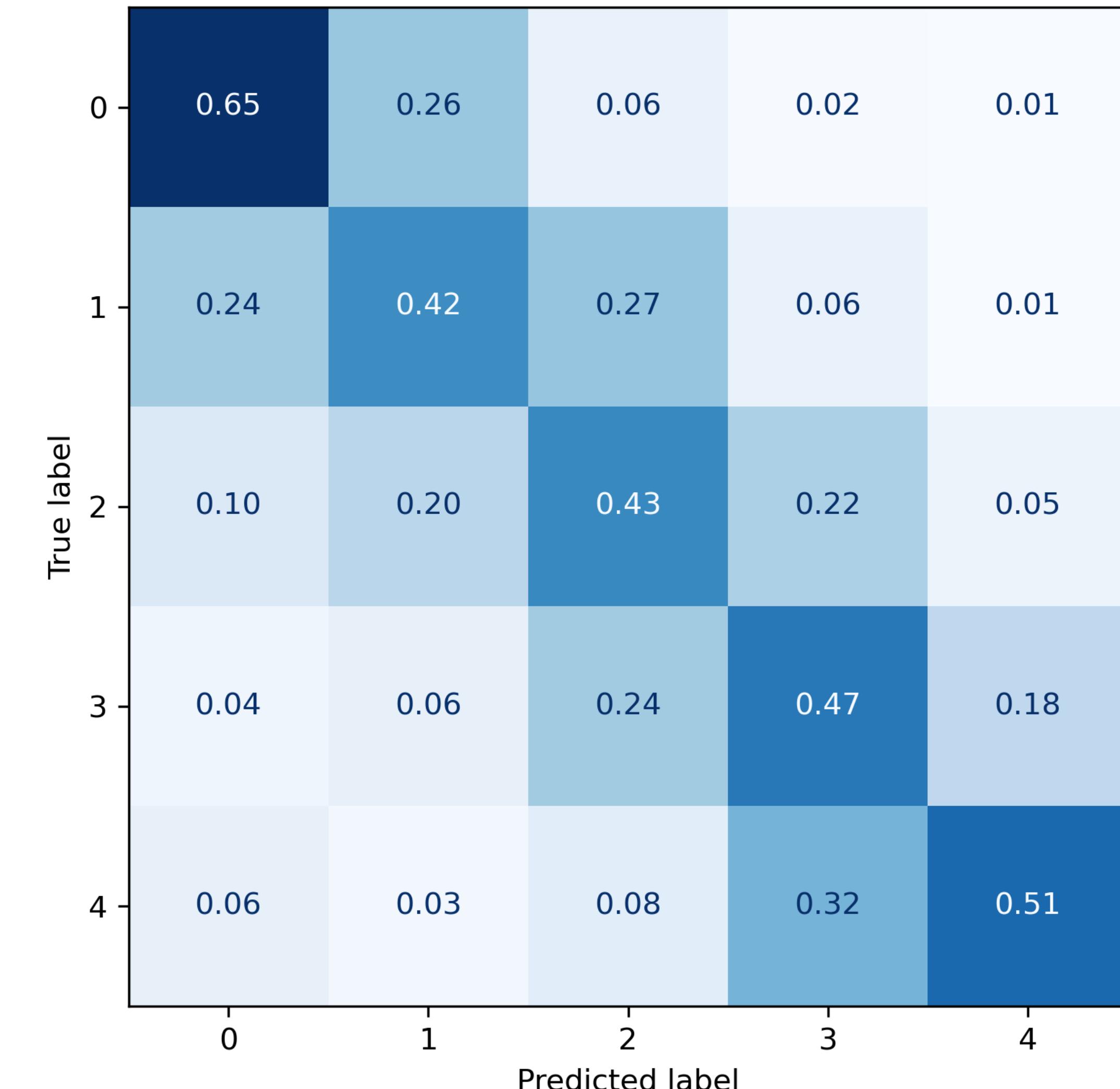
Model Rating Prediction: 2

NaiveBayes Classifier Test Confusion Matrix

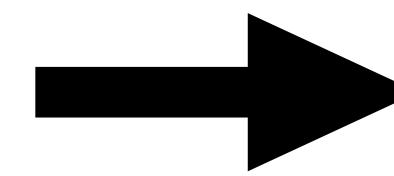
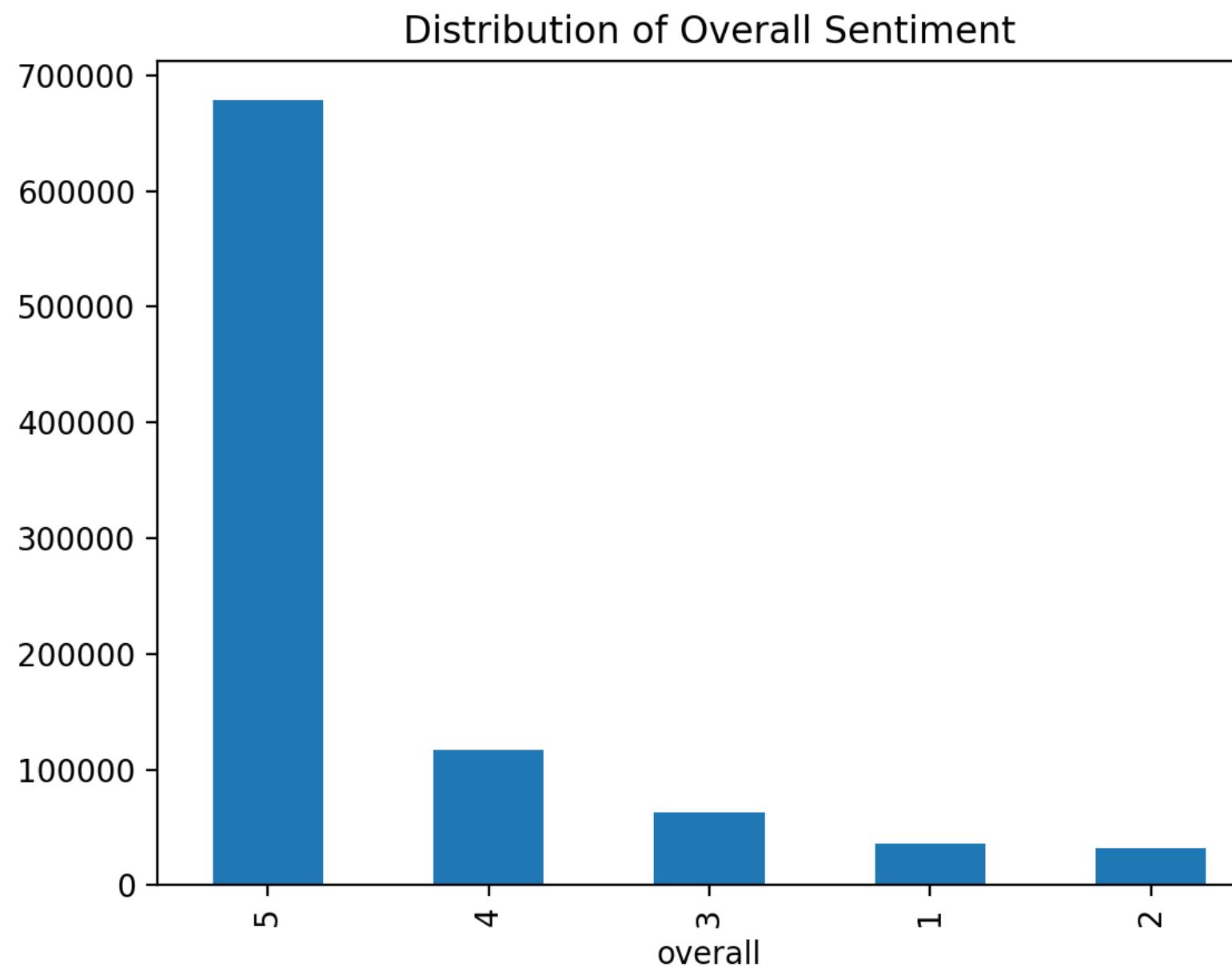
Normalized Confusion Matrix Test Set: BoW



Normalized Confusion Matrix Test Set: TF-IDF



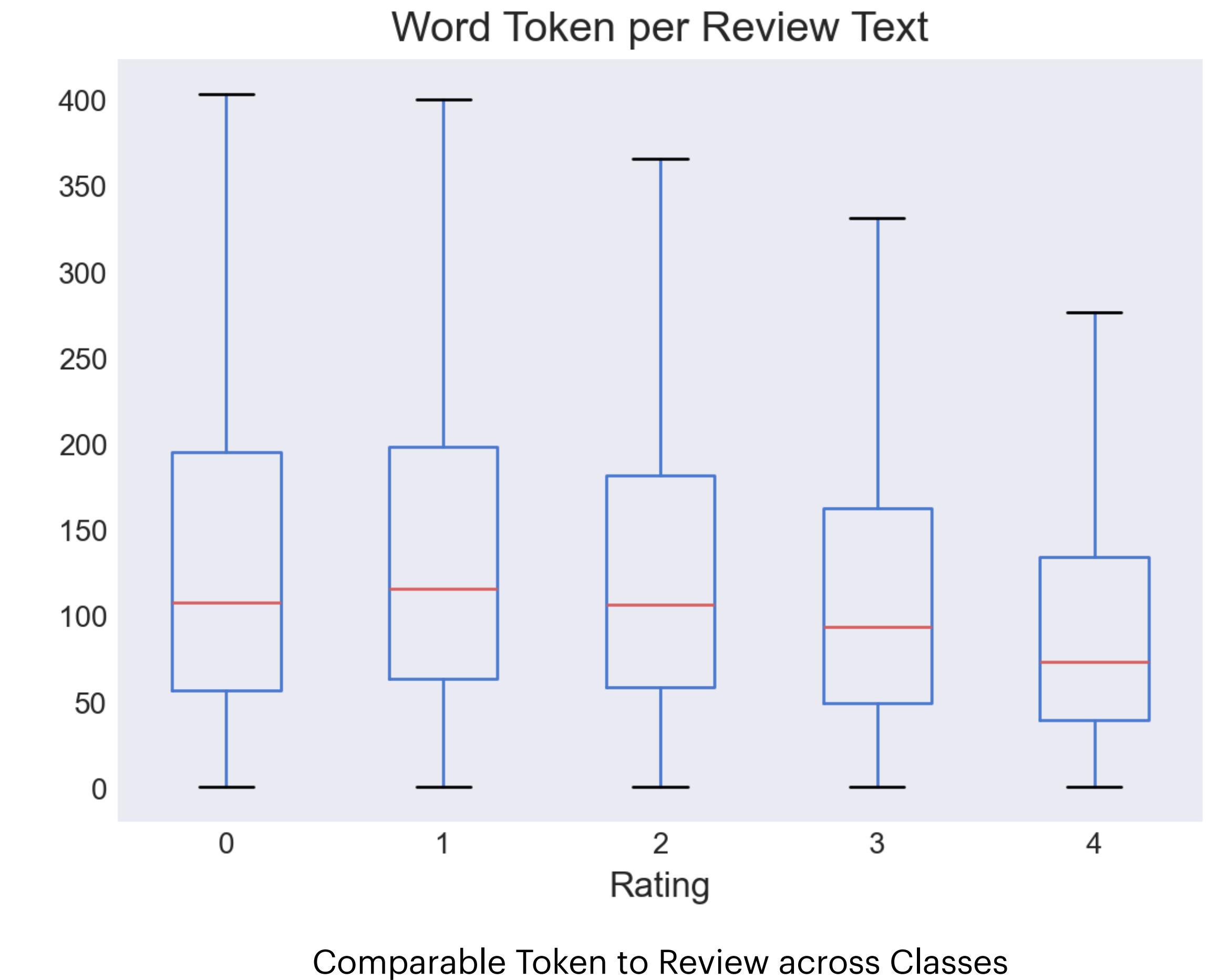
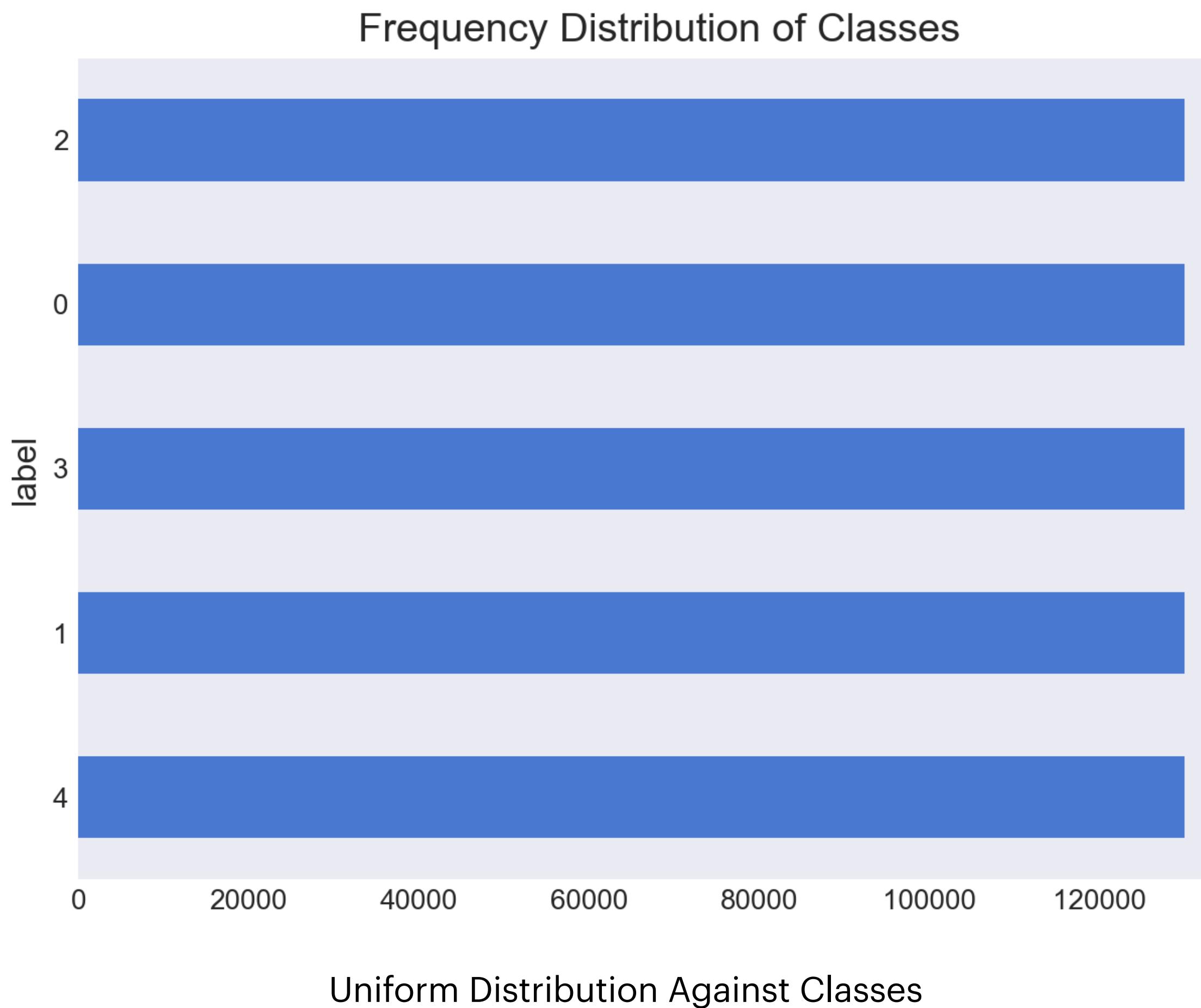
Challenge I: Majority Class Problem



Sample WordCloud: Positive contexts dominates vocabulary



Solution I: Switching to Yelp Dataset

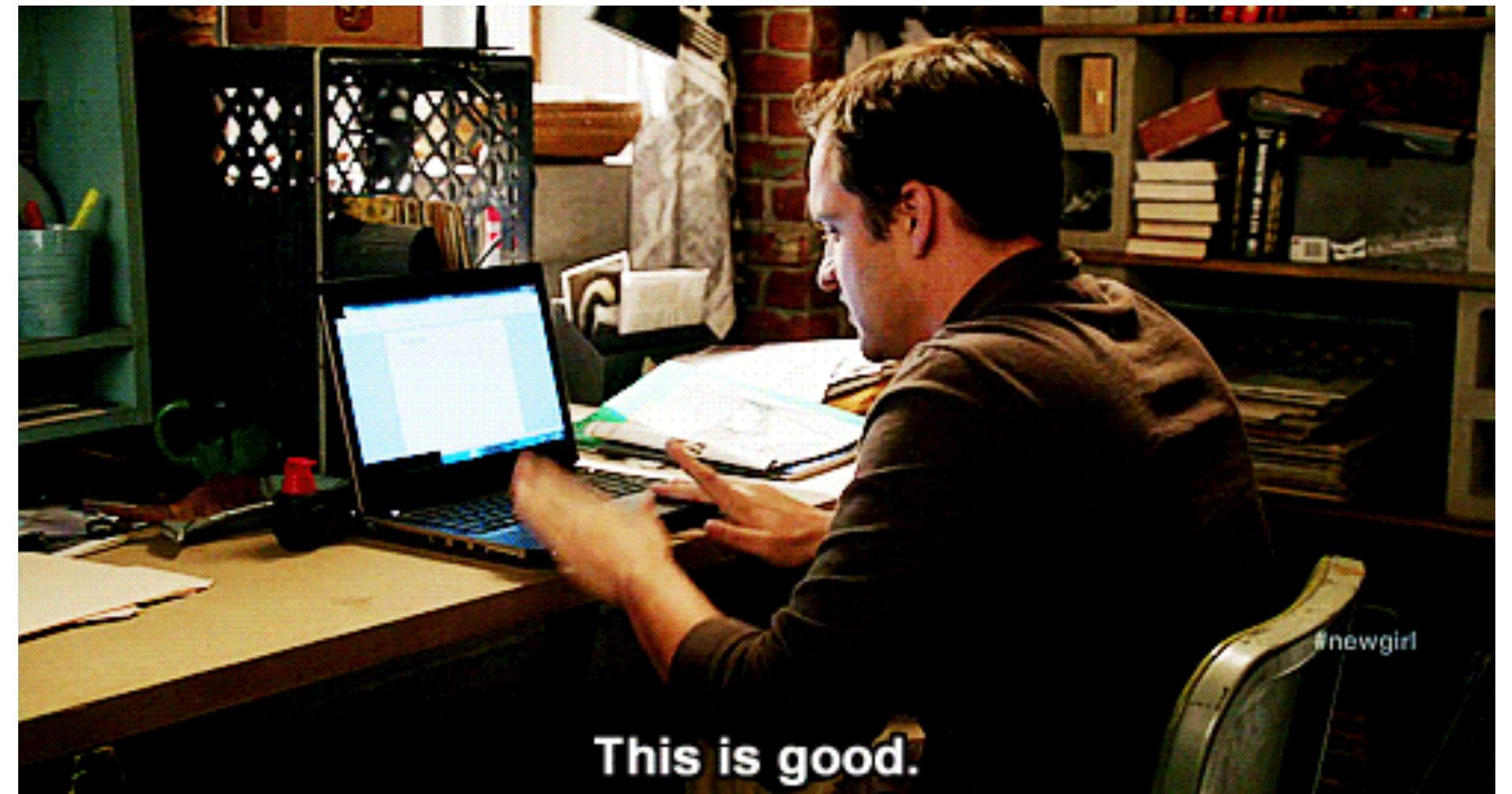


Challenge II: Computational Limitations

```
def preprocess(text):  
  
    # replacing & with and  
    text = re.sub("[^a-zA-Z]", " ", text.replace('&', 'and').lower())  
    text = replace_digits_with_text(text)  
    text = lemmatize_text(text)  
    return text
```

```
[2]: def preprocess(text):  
  
    # Replacing & with and  
    text = re.sub("[^a-zA-Z]", " ", text.replace('&', 'and').lower())  
    text = replace_digits_with_text(text)  
    text = lemmatize_text(text)  
    return text  
  
[*]: train_data = pd.read_csv('yelp_train_df.txt')  
train_data.loc[:, 'clean_text'] = train_data['text'].apply(lambda x: preprocess(str(x)))
```

2 Hours Later



This is good.



OURGANGAPPRECIATION

Challenge II: Computational Limitations

```
[2]: def preprocess(text):

    # Replacing & with and
    text = re.sub("[^a-zA-Z]", " ", text.replace('&', 'and').lower())
    text = replace_digits_with_text(text)
    text = lemmatize_text(text)
    return text

[*]: train_data = pd.read_csv('yelp_train_df.txt')
      train_data.loc[:, 'clean_text'] = train_data['text'].apply(lambda x: preprocess(str(x)))
```

4 Hours Later

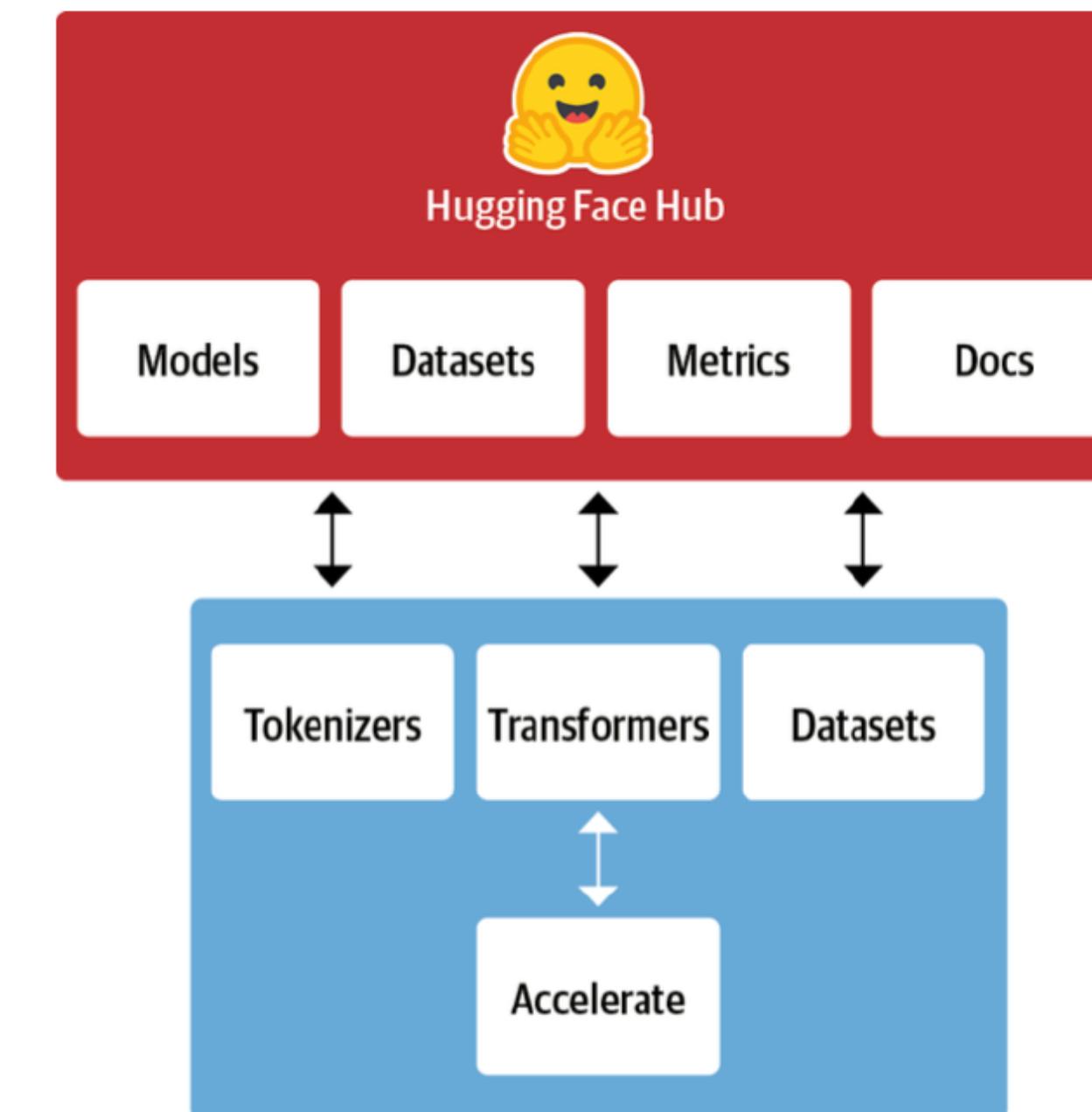


Solution II: Switching Tools for Performance Improvement

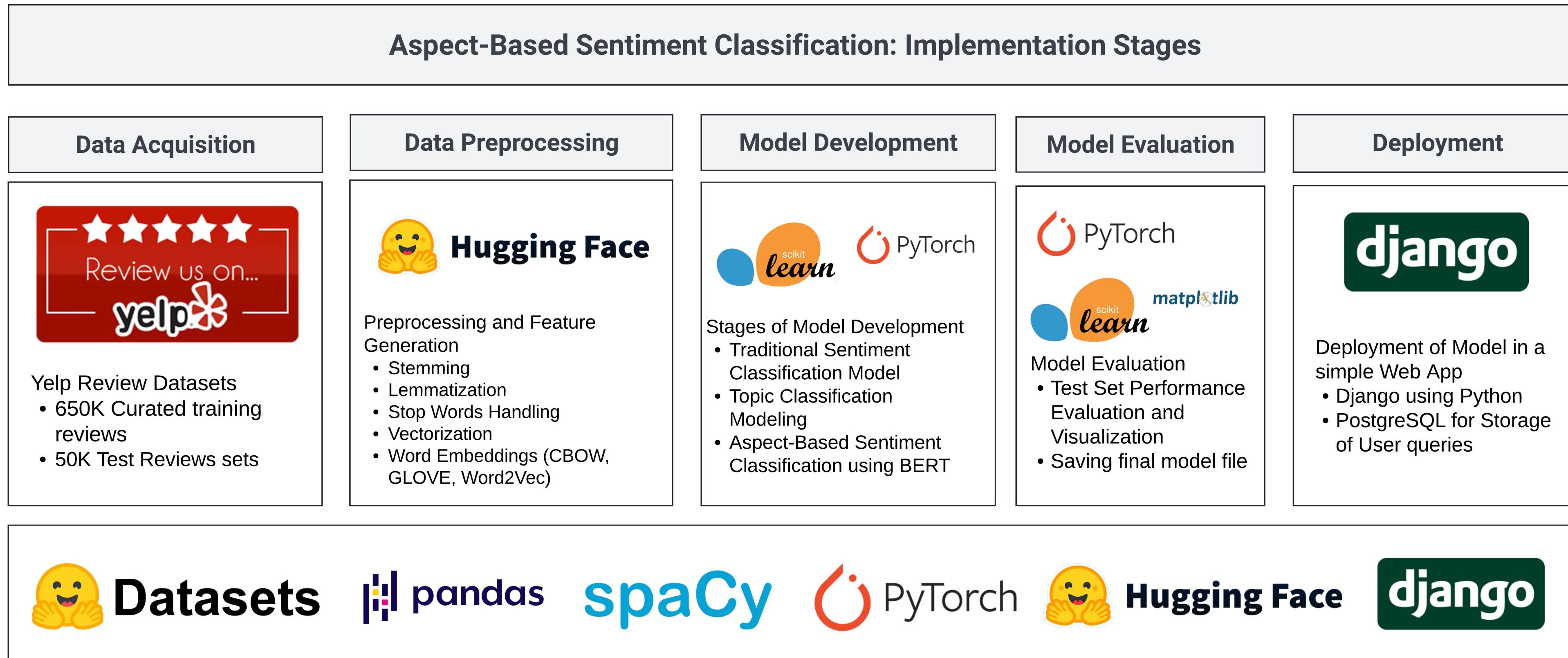
Experimentation Tools



Adopting Efficient Tools



Revised Implementation Plan



Resources and Links:

Yelp Review Dataset: <https://www.yelp.com/datasets>

HuggingFace Datasets: <https://huggingface.co/docs/datasets/en/index>

Spacy Package: <https://spacy.io/>

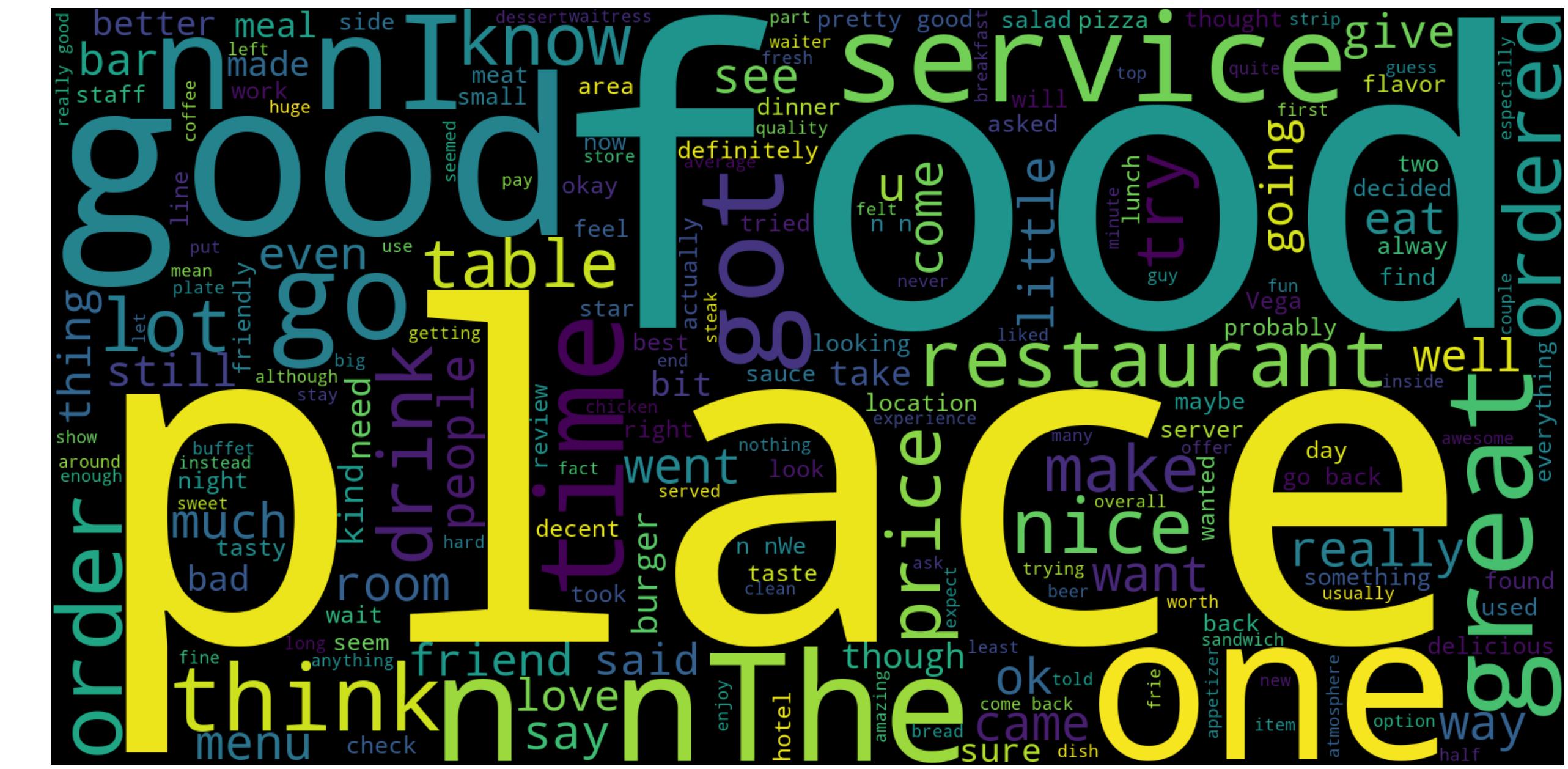
Spacy NLP Tutorial: <http://course.spacy.io/>

Yelp Dataset: WordCloud by Rating

Class Labels: 0 and 1 - Negative



Class Labels: 2 - Neutral



Yelp Dataset: WordCloud by Rating

Class Label: 3 - 4 - Positiv

