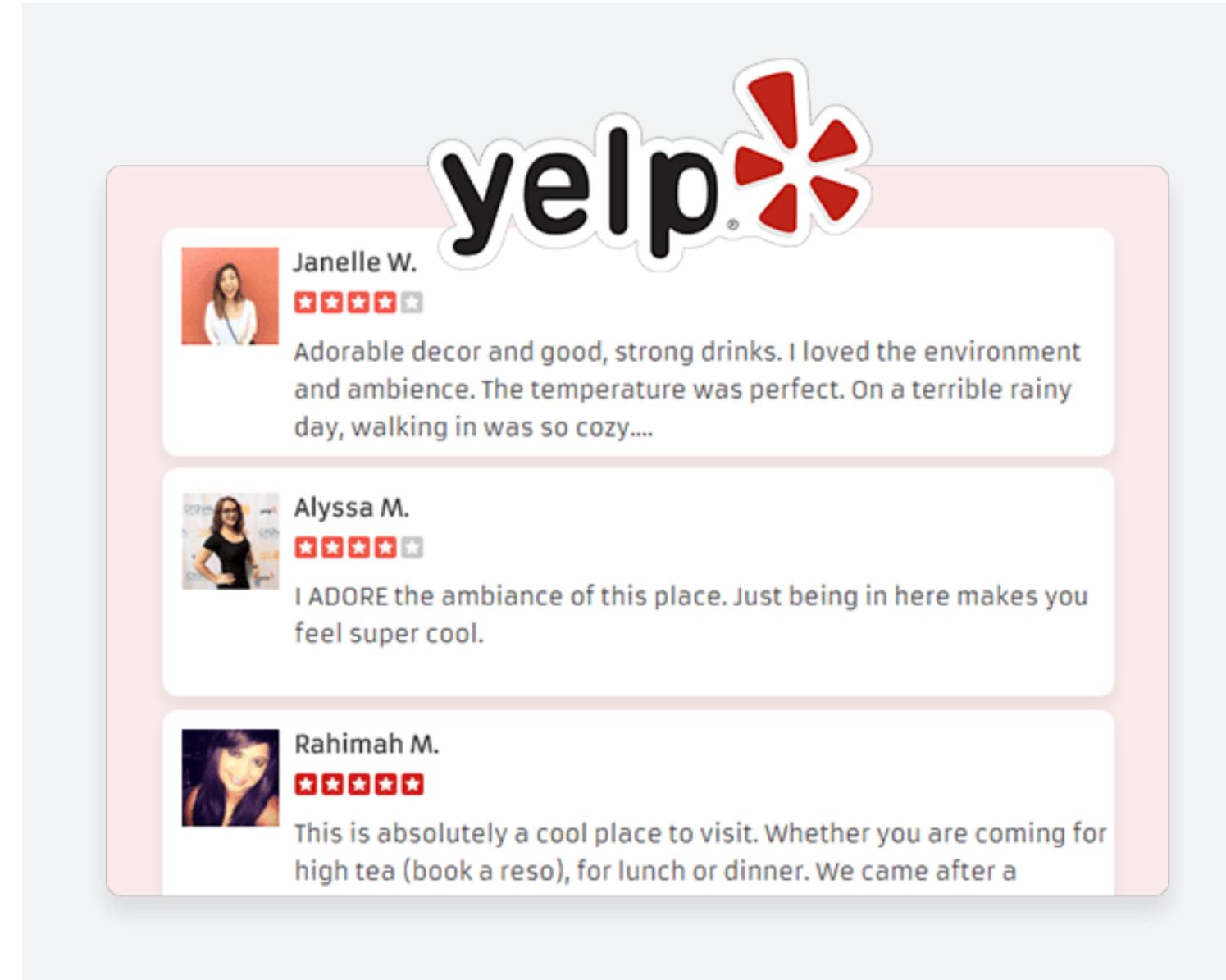


Aspect-Based Sentiment Analysis

Final Presentation



Data Mining - Spring 2024

Sifael Ndandala, Protogene Hahirwabayo, Piumi Abeyrathne, 14. Mai. 2024

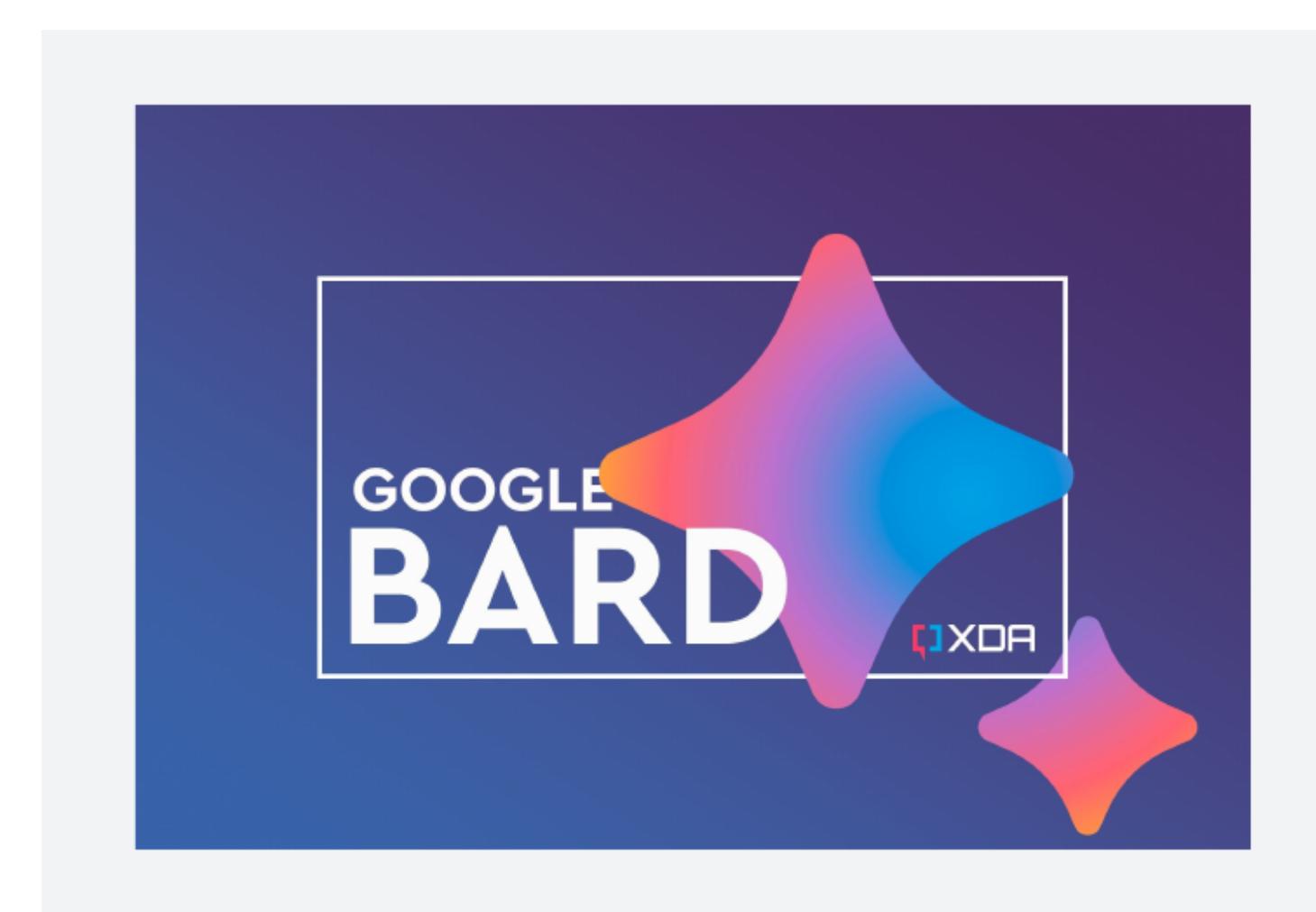
Goals and Motivation

Adoption of LLM Generative AI Technology across industries



ChatGPT

Opportunity in Natural Language Processing to develop fundamental technical skills and technology stack



Real World Application: Using Reviews to understand and provide insights that help businesses

Project Overview

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

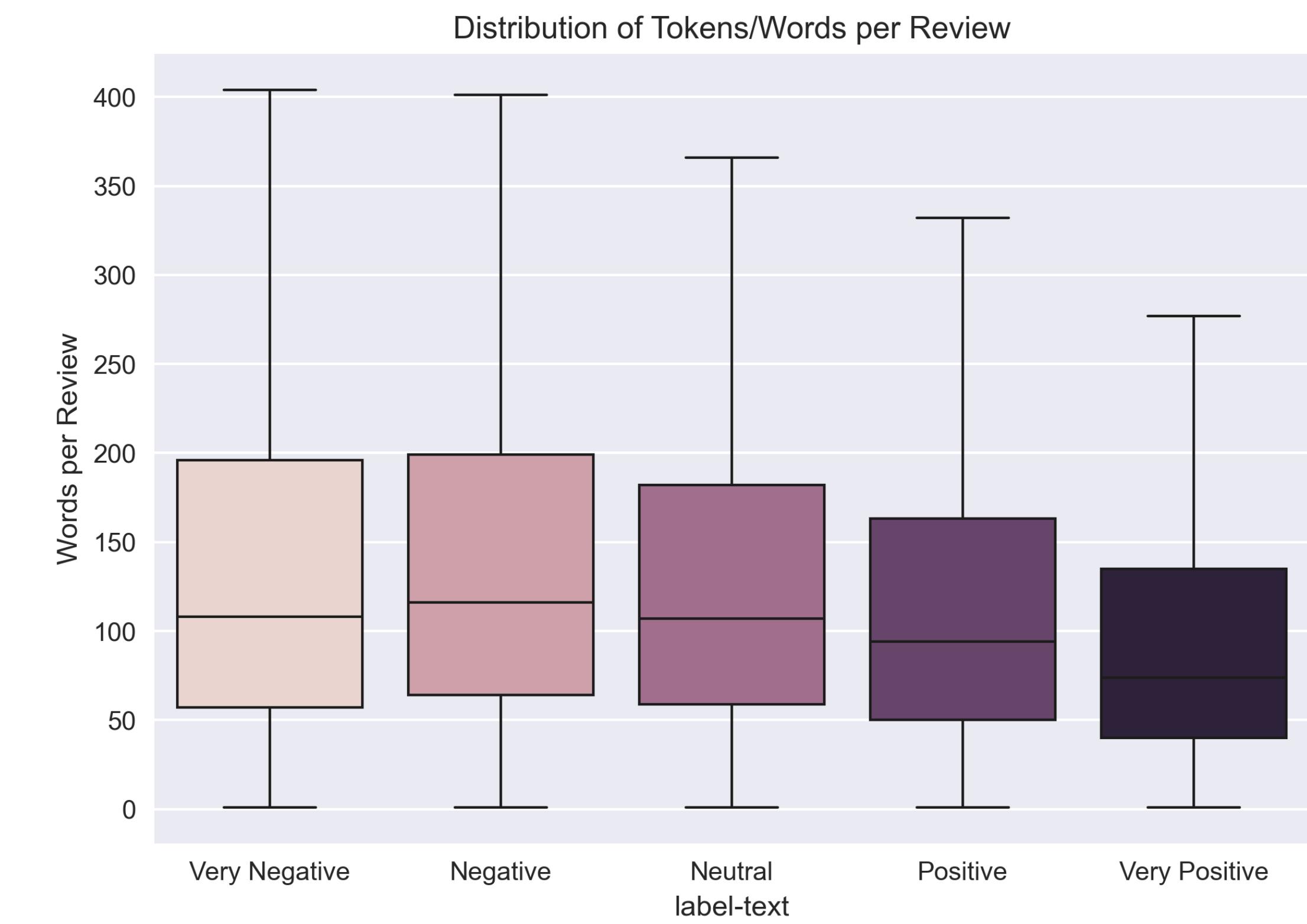
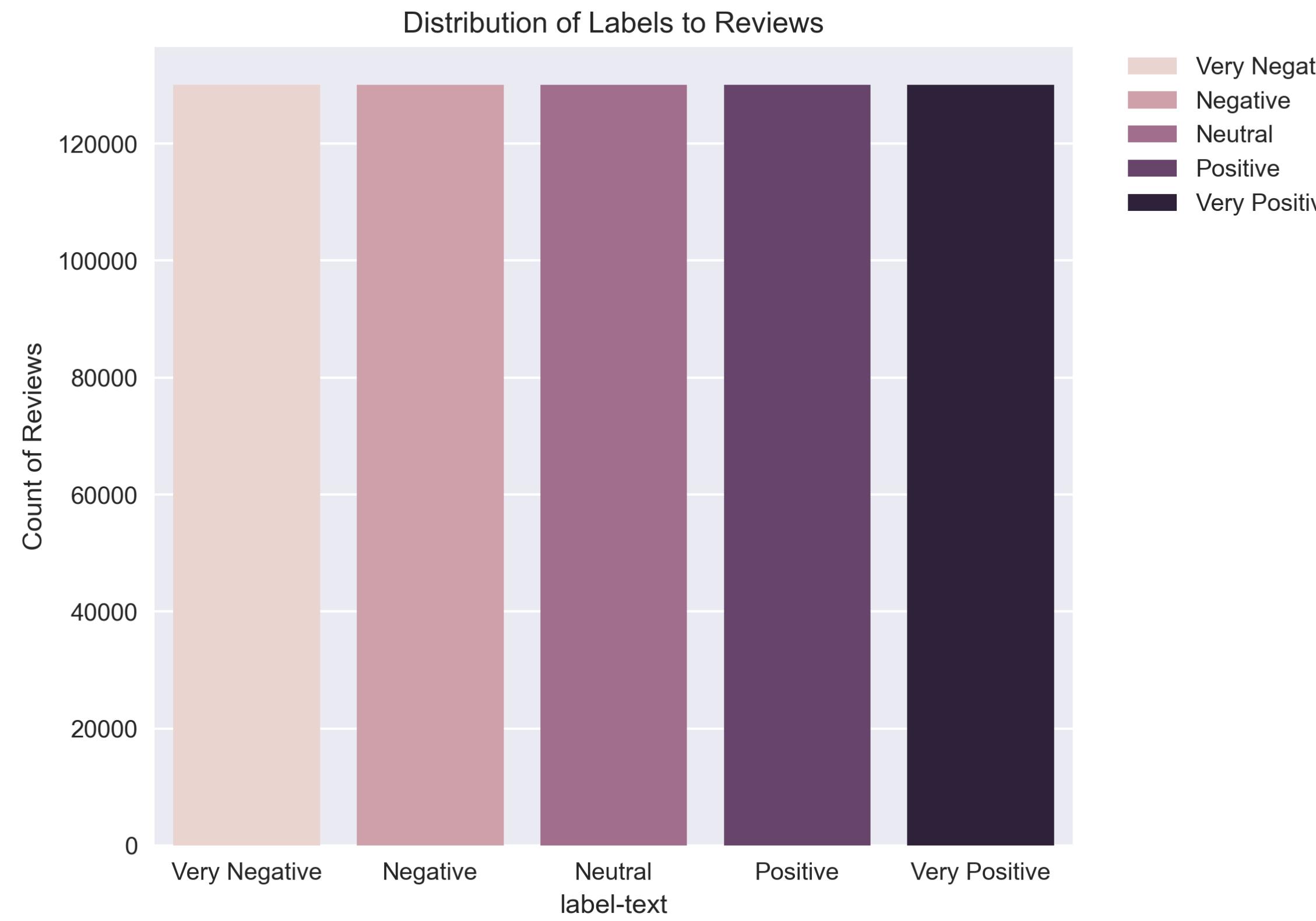


Wifi



The Yelp Dataset

Train set: 650,000 Reviews with 5 Categories, Test Set: 50,000 Reviews with 5 Categories



Sample Reviews

```
train_df = pd.read_pickle('train_df.pkl')
train_df[['text', 'label-text']].sample(5).style.hide()
```

text	label-text
Such a cute place - too bad we were the only two people there on a Tuesday night at 11:00 pm when we'd made the trip all the way down there from the Strip and it doesn't open until 9:00 and the door is locked until we kill some time waiting for it to \"happen\". It never did. It's very pink and sparkly though.	Negative
We had a reservation for two at 7:00pm, got to the restaurant and our table was not ready... They told us the people before were still eating!! We finally got our table at 7:15, our waiter came to our table not before 7:45-8:00 to get our order. I ordered a drink and I had to remind her after half an hour about my drink. Got our appetizer by 8:45, the waiter made a mistake in our order! By 9:00 (2 hours after we got here) we finally got our sushis. The Wasabi & Ginger next to our sushi had changed their colors, soooo dry that we couldnt seperate the layer of ginger... Sushi was ok, you can get the same sushi or even better ones at any \"Sushi Shop\" or \"Sushi man\" take out place. I rate them 1, because they have potential... Its a nice place but food and service Not so great.	Very Negative
I'm originally from the MidWest so this place reminds me of home...cowboy hats, mason jars, country music, the electric slide...as far as a country bar goes...you really can't ask for much more! \n\nI haven't actually been the bar in probably 3 years...they have added stripper poles with \"dancers\". Very pretty girls but a little random. People are eating dinner and there's this half naked girl dancing to some country song! Probably not the best place to take small children for dinner! \n\nWe went there because my husband's best friend had Navy friends in town. They said they liked the food...didn't see what they had. \nService is okay...They were busy and don't have very many people working at the bar. You're going to have to wait to get a drink if it's at all busy in there. Drink prices are pretty typical for a Strip eatery/bar. \n\nMy favorite thing about going here is you have choices in a close proximity. Carnival court, casino gambling, etc. are all pretty close in walking distance.	Neutral
No this is not a lesbian bar or a strip club.. just a gay bar.. no attitudes and a very comfortable space to hang out in! Plus they have Blue Moon on tap which is fantastic!! \n\nMe and the Mr rarely go to the bars because they can be so cruisy, not this place. Just a local drinking spot for gay men, oh yes, and there are loads of bears too!	Positive
Bone-in ribeye - a bit tough, Outback/Saltgrass quality. \n\nAsparagus - decent seasoning, unspectacular. \n\nFried mac n' cheese - the Texas State Fair could beat it out. \n\nMashed potatoes (menu calls it something else) - best item ordered, too bad it was one fifth of the proper serving. \n\nNice server. Food came as slow as molasses. iPad drink menus. Whoopie. \n\nFour stars if 15 – 20/head, I paid 65 with one drink. Not a happy camper for value received. As a douchey doorman once told me, welcome to Vegas!	Negative

Random Sample of 5 from train dataset

Breaking Down the Problem

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.

Identify Aspects

Hotel service

Food (breakfast)

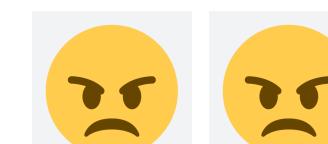
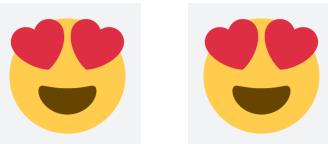
Wifi/Internet

Classify Aspect Sentiment

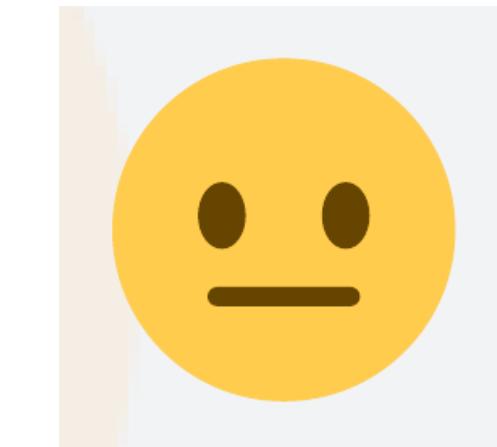
Hotel service

Food (breakfast)

Wifi/Internet



Classify Overall Sentiment



Task I: Aspect Extraction with Structural Dependencies

What is Dependency Parsing?

The process to analyze the grammatical structure in a sentence and find out related words as well as the type of the relationship between them.

Sample sentence: “*The breakfast was amazing with lots of variety*”

Token	Relation	Head	Children
The	det	breakfast	[]
breakfast	nsubj	was	[The]
was	ROOT	was	[breakfast, amazing]
amazing	acomp	was	[with]
with	prep	amazing	[lots]
lots	pobj	with	[of]
of	prep	lots	[variety]
variety	pobj	of	[]

The tokens and their parts of speech are:

The	DET
breakfast	NOUN
was	AUX
amazing	ADJ
with	ADP
lots	NOUN
of	ADP
variety	NOUN

Task I: Our Implementation

Example: However the food was terrible. Breakfast ran from 7-9 am and there was little variety.

Step 1: Tokenization and POS Tagging completed.

Tokenized and tagged sentence: [('However', 'RB'), ('the', 'DT'), ('food', 'NN'), ('was', 'VBD'), ('terrible', 'JJ'), ('.', '.'), ('Breakfast', 'NNP'), ('ran', 'VBD'), ('from', 'IN'), ('7-9', 'JJ'), ('am', 'VBP'), ('and', 'CC'), ('there', 'EX'), ('was', 'VBD'), ('little', 'JJ'), ('variety', 'NN'), ('.', '.')]

Step 2: Combining consecutive nouns completed.

New word list after combining nouns: ['However', 'the', 'food', 'was', 'terrible', '.', 'Breakfast', 'ran', 'from', '7-9', 'am', 'and', 'there', 'was', 'little', 'variety', '.']

Step 3: Filtering stop words completed.

Words after stop word filtering: ['However', 'food', 'terrible', '.', 'Breakfast', 'ran', '7-9', 'little', 'variety', '.']

Step 4: Dependency Parsing

Dependency parse results: [('However', 2, 'advmod'), ('food', 0, 'root'), ('terrible', 2, 'amod'), ('.', 2, 'punct')]

Step 5: Extracting Dependency Features

Feature: {'However': 'RB', 'food': 'NN', 'terrible': 'JJ', '7-9': 'JJ', 'little': 'JJ', 'variety': 'NN'}

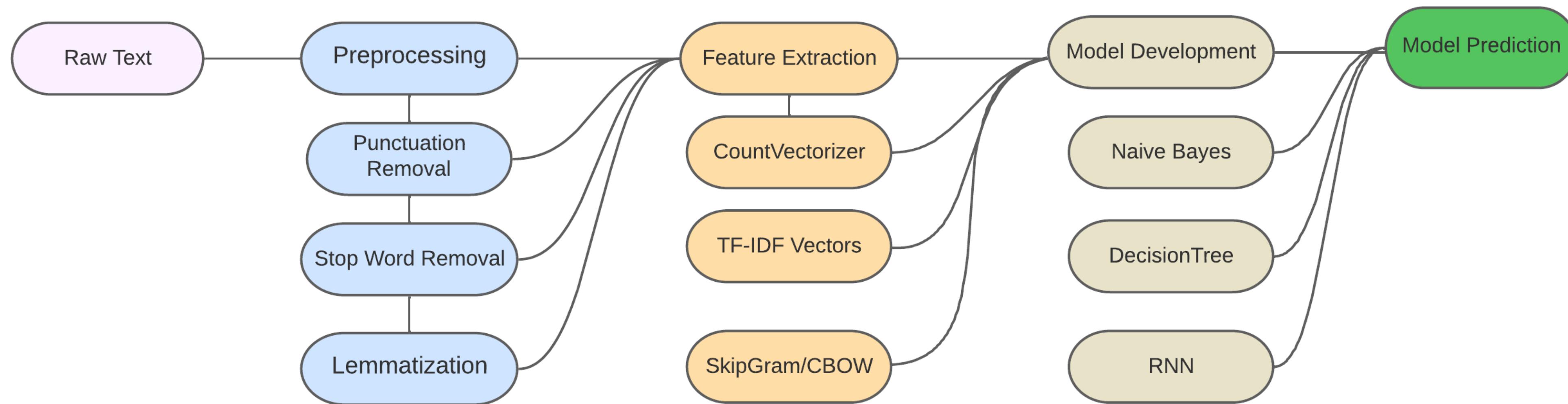
Step 6: Extract Aspects

Final Aspects: ['food', 'variety']

['food', 'variety']

Task II & III: Sentiment Analysis

Sentiment Classification Pipeline



Sentiment Analysis Implementation

User Input Text:

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.

Preprocessing:

[punctuation removal, tokenization, and lemmatization]

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

Feature Generation:

[1-gram/2-gram model]

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

Vector Size: 14102

Sentiment Model Prediction:

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

Model Rating Prediction: 2

Sentiment Analysis Model Results

Features Extractions Methods		Naive Bayes		Decision Trees	
		Train	Test	Train	Test
1-gram	Count Vectors	65	49	100	35
	TF-IDF	68	49	100	35
2-gram	Count Vectors	83	45	99	35
	TF-IDF	86	46	99	34

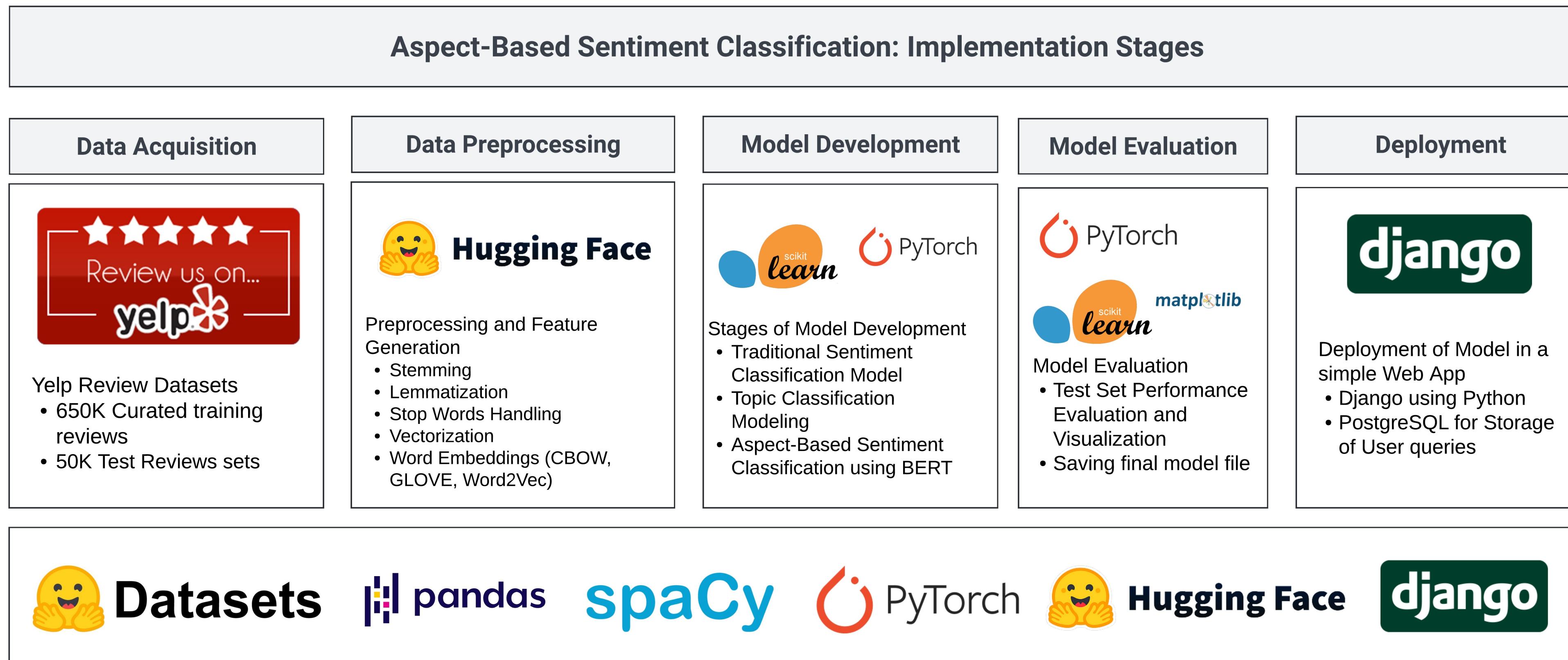
Challenges and Lessons Learned

Parallel and Distributed Computing is an exceptionally important toolkit

Stand on top of giants. Use pre-trained models whenever possible

Develop ML training pipelines and processes based on check-points

Overall Implementation Steps



References

https://downloads.cs.stanford.edu/nlp/software/dependencies_manual.pdf

https://lg845.github.io/LAEL_CoreNLP/

<https://universaldependencies.org/u/overview/syntax.html>