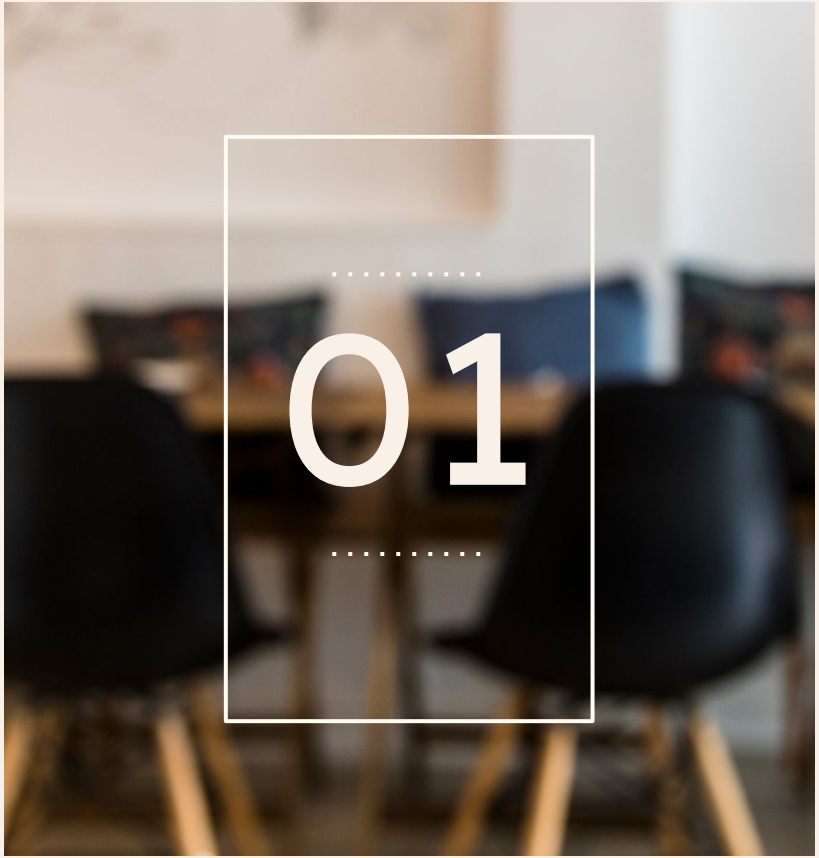
Neutral Review Classification

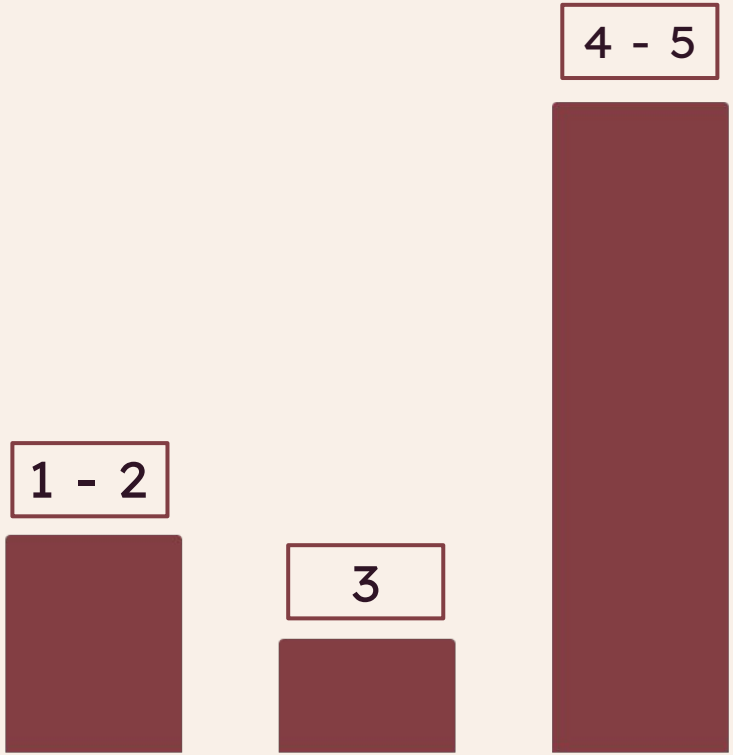
Victor Chen



01

Business Problem

The Neutral Feedback Problem



1 - 2

Improvements

1 - 2 star reviews typically indicate things that need to be improved

3

N/A

Neutral reviews cannot provide informational value unless a human interprets the context

4 - 5

Highlights

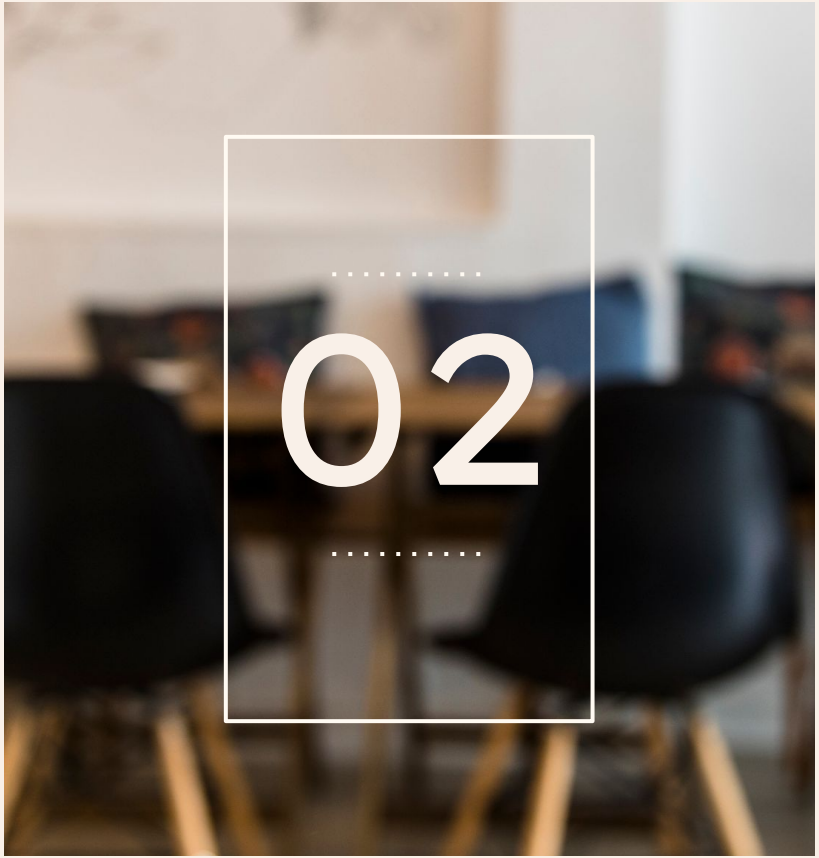
4 - 5 star reviews typically indicate highlights that made the experience worthwhile

BUSINESS PROBLEM

Small businesses, such as the tea/coffee shop Boba Drive, often suffer from a lack of business intelligence

Goal:

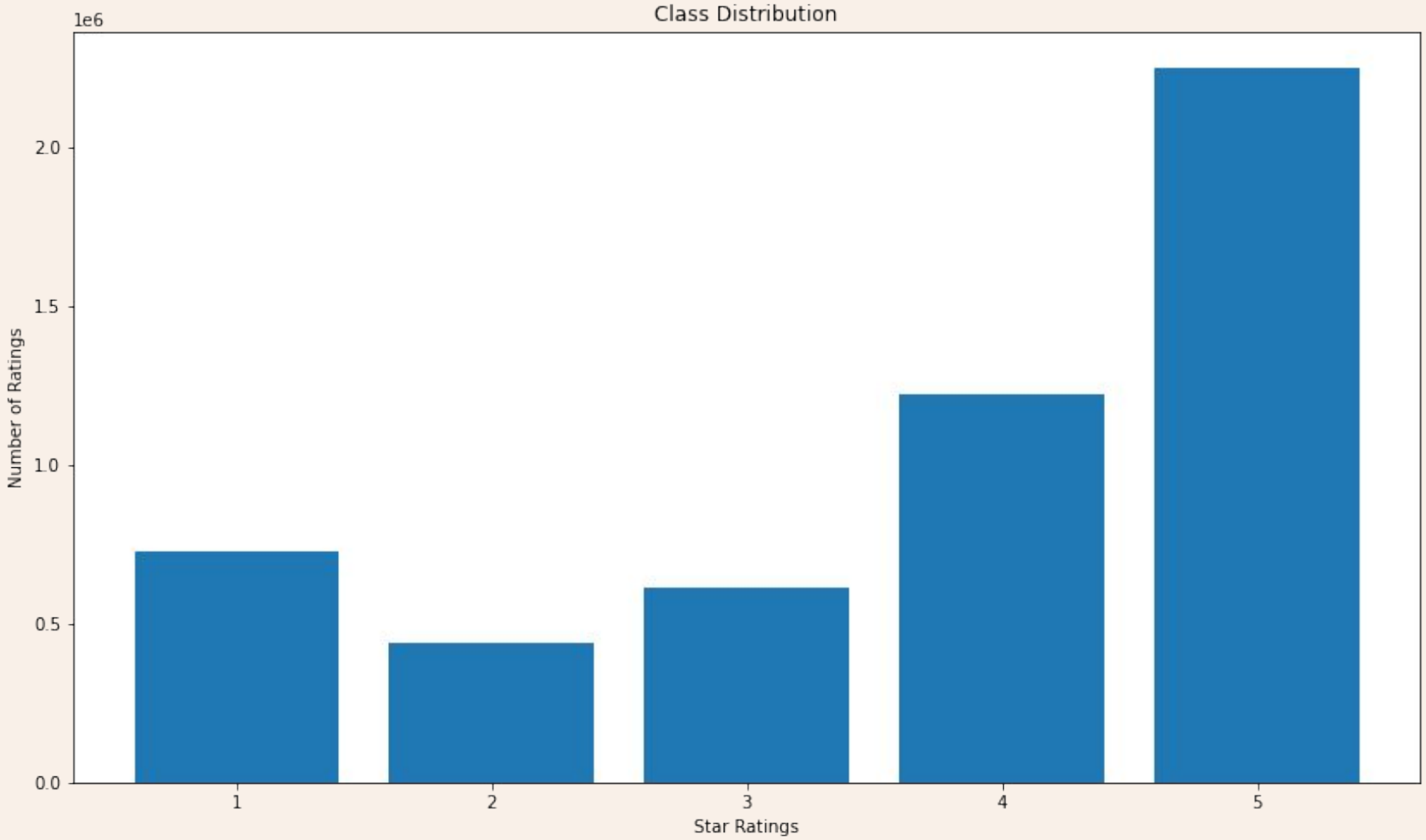
Add more value to difficult-to-interpret reviews (neutral reviews) by classifying them as positive or negative using NLP



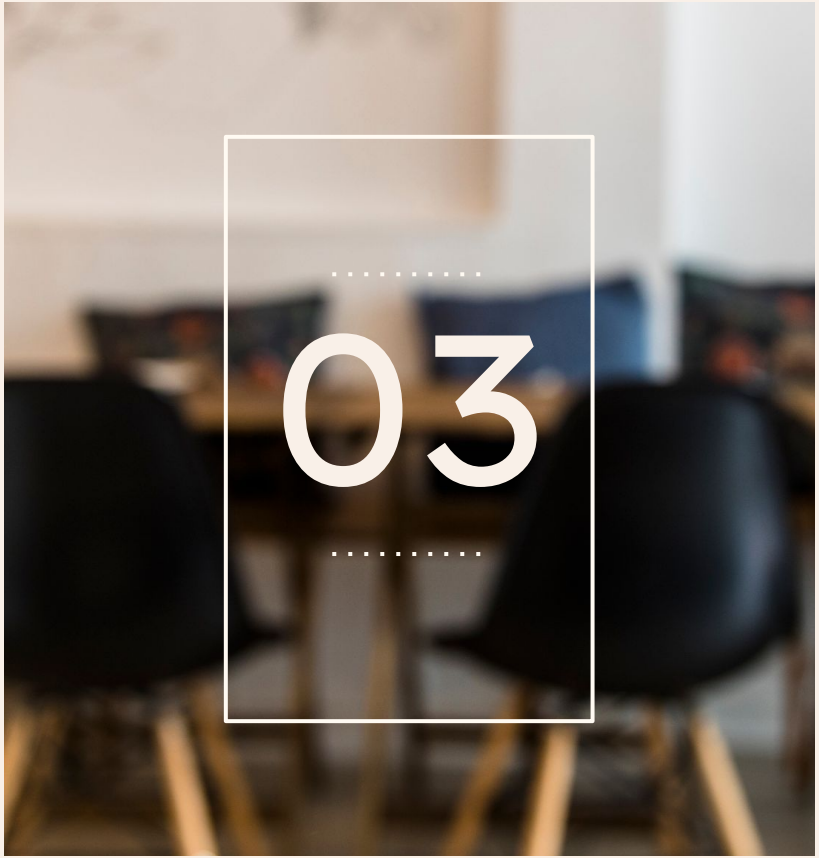
Data

Yelp Reviews Dataset

- From: Kaggle
- **5,261,668 reviews**
- All different domains, including restaurants, bars, hotels, etc.
- Majority of this project sampled 500,000 reviews from this dataset


Class Distribution
1e6
2.0
1.5
1.0
0.5
0.0
Number of Ratings
1
2
3
4
5
Star Ratings

Word Cloud



Modeling

Baseline Models (Non-neutrals)

Multinomial Naive Bayes

Training F1: 0.926
Testing F1: 0.922

Random Forest

Training F1: 1.000
Testing F1: 0.922

Logistic Regression

Training F1: 0.974
Testing F1: 0.970

Modeling Process

Solver

From lbfgs to newton-cg

Max Iterations

From 100 to 500

C (regularization strength)

From 1 to 3

Final Model (Non-neutrals)

Logistic Regression

Training F1: 0.977
Testing F1: 0.971

Predicting Neutral Reviews

Baseline Model

Validation F1: 0.680

Final Model

Validation F1: 0.687

The neutral review classification training and testing data was classified manually by me. This most definitely leads to a level of bias no matter how hard I tried to remain impartial.

Bias Disclaimer

Next Steps

More Neutral Reviews

The more neutral reviews are manually classified, the better the model can train itself

Scaling for Domain Expertise

Reviews are all over the place right now. Domain specific models will be helpful

Key Words

Find most weighty words for both negative and positive reviews in each domain

Web App

Ability to copy paste in a review, and get the classification of ‘Negative’ or ‘Positive’ along with feature importance of words

THANKS

DO YOU HAVE ANY
QUESTIONS?

victor.i.chen.98@gmail.com
https://github.com/vchen-98
https://www.linkedin.com/in/victorchen98/