



Sentiment Analysis of Interview Reviews

John Lim
DSI 17

Problem Statement

- A company has approached my team to review their interview process.
- The client wants to give their candidates the best interview experience regardless of whether they were accepted for the job.
- By doing so, they hope to achieve more positive interview reviews, strengthen the company branding and ultimately attract more talent.



INTRO COMPANY WHAT TO EXPECT SETUP PRACTICE INTERVIEW

Congratulations! You have completed your digital interview.

Continue

Feedback

We are continuously improving our platform and appreciate your feedback!

Submit

Content



METHODOLOGY



DATA
GATHERING



EDA AND PRE-
PROCESSING



MODELLING



FINDINGS AND
LIMITATIONS



Methodology

- Analysing companies with thousands of interview reviews on glassdoor
- Build a model that is optimized and tuned based on accuracy score.
- The model will be able to tell if an interview feedback is a positive or negative sentiment
- Companies can then use the insights to train their HR and hiring managers to adopt or avoid certain attributes during an interview process





Data Gathering...

Web Scraping glassdoor
interview reviews with
python, selenium and
headless chrome



Software Engineer - New Grad Interview

Anonymous Interview Candidate in Seattle, WA (US)

☐ No Offer

☐ Negative Experience

☒ Easy Interview

Application

I applied online. I interviewed at Microsoft (Seattle, WA (US)) in October 2020.

Interview

I was emailed by a recruiter around 2 weeks after applying for a first interview. That went pretty well, only one LC easy and background questions. I was called back for a second round which was supposed to be a day with 4 interviews. They canceled the last one and told me only after I emailed them during the supposed time. The three other interviews were pretty bad, interviewers didn't have their cameras on and didn't seem interested in what I have to say at all. I also haven't heard back from them yet.

Interview Questions

Top 5 numbers in an array without sorting, Remove duplicates from array in place, Binary search algo, Valid parans
What are distributed systems, microservices, APIs
What do you think about Agile development

[Answer Question](#)





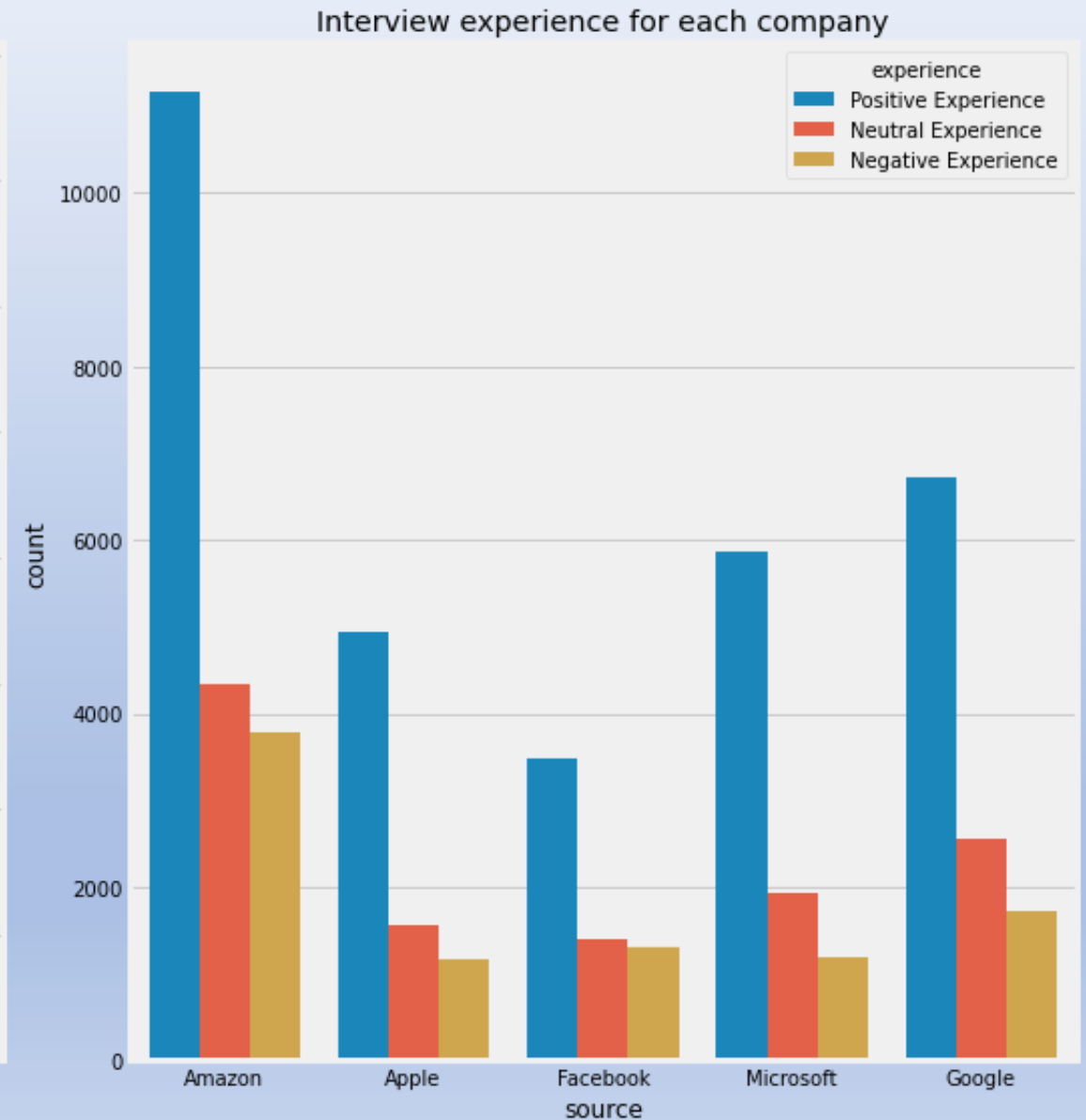
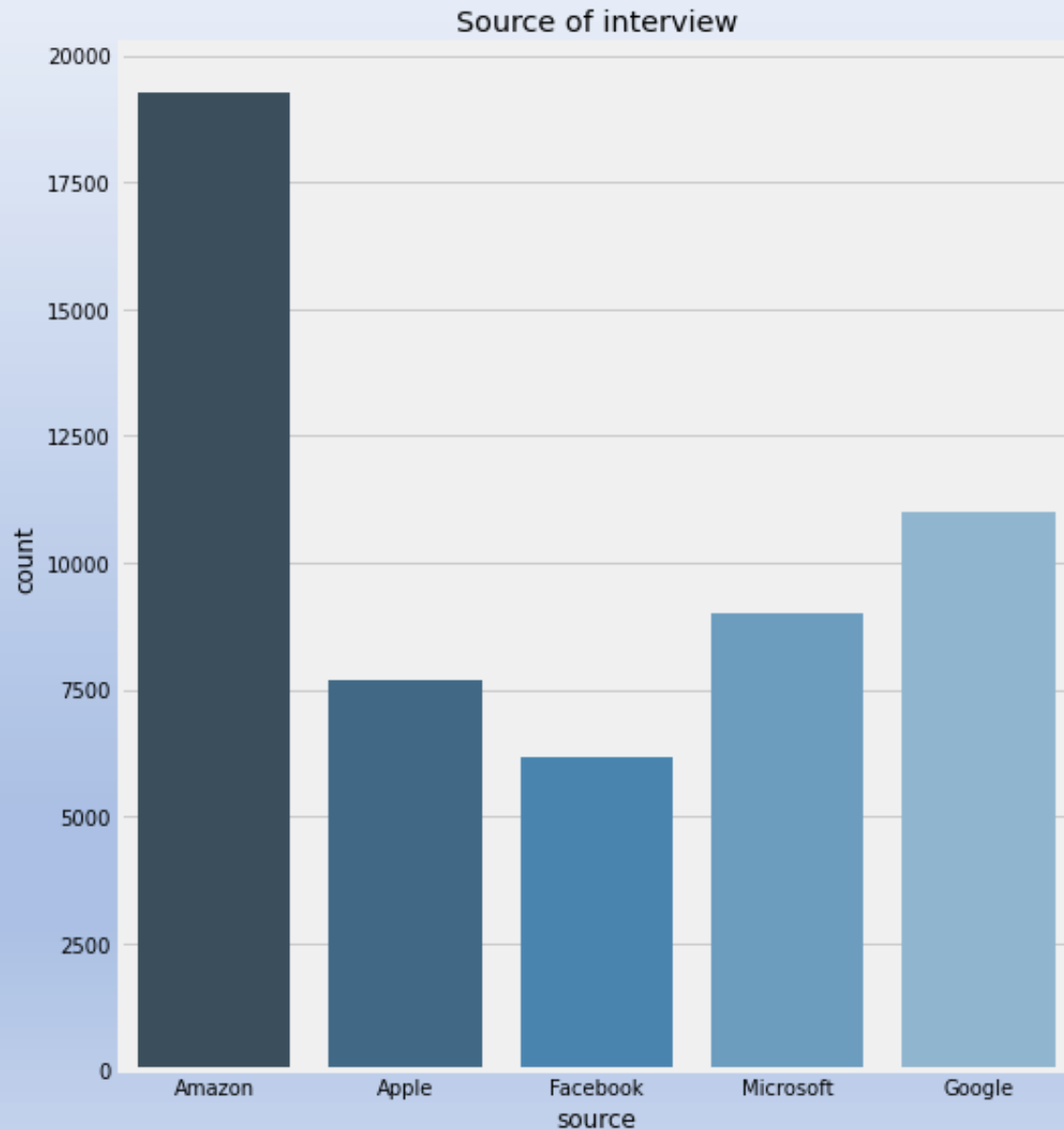
Data Gathering



Web Scraping Apple, Microsoft, Facebook, Google and Amazon profiles on glassdoor for their interview reviews

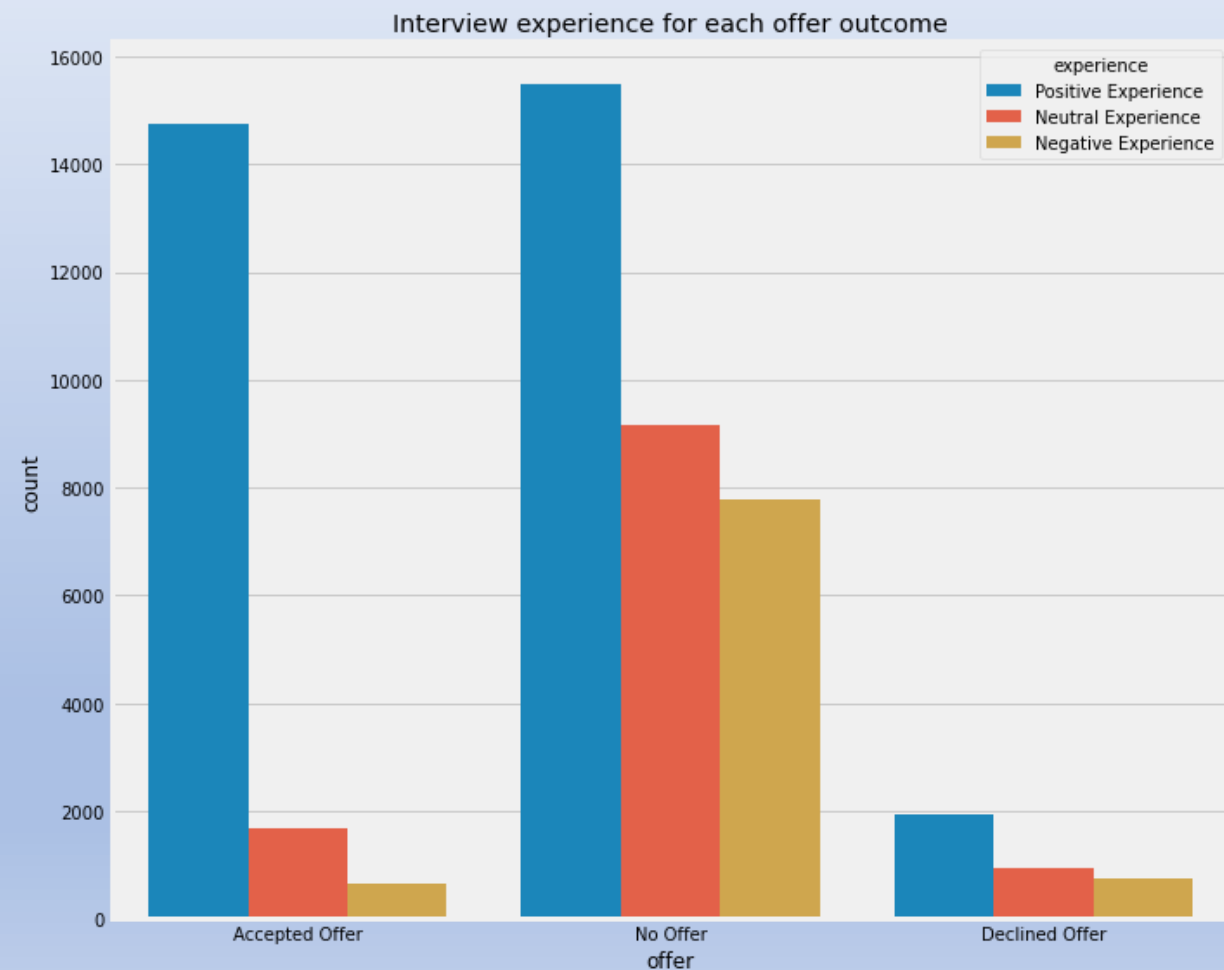
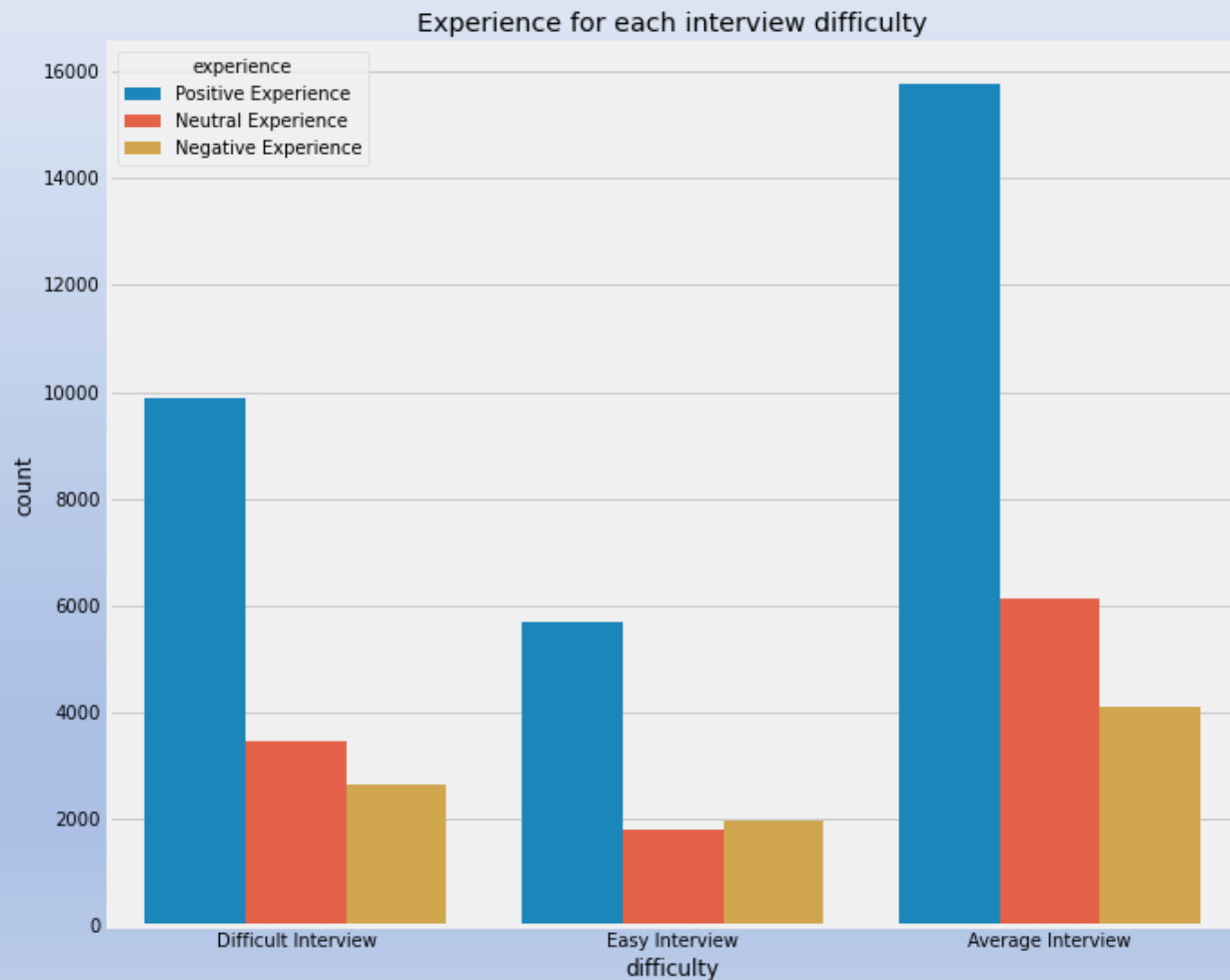


Interview experience of different companies





Experience in each Interview difficulty/acceptance



Drop neutral experience

- Adopt attributes that lead to positive experience
- Avoid attributes that lead to negative experience
- Many of the neutral experience is might not be neutral in fact



Procurement Analyst Interview

Anonymous Interview Candidate

☐ No Offer

☒ Neutral Experience

☐ Difficult Interview

Interview

A phone interview was conducted by someone in the same position at a different location. He described the role and didn't seem too interested in my response. This was surprising considering the size of the company. I'd expected the interview questions to be more professional and less chauvinistic.

Interview Questions

You do know this isn't a job that requires you to wear high heels, right?

↳ 1 Answer



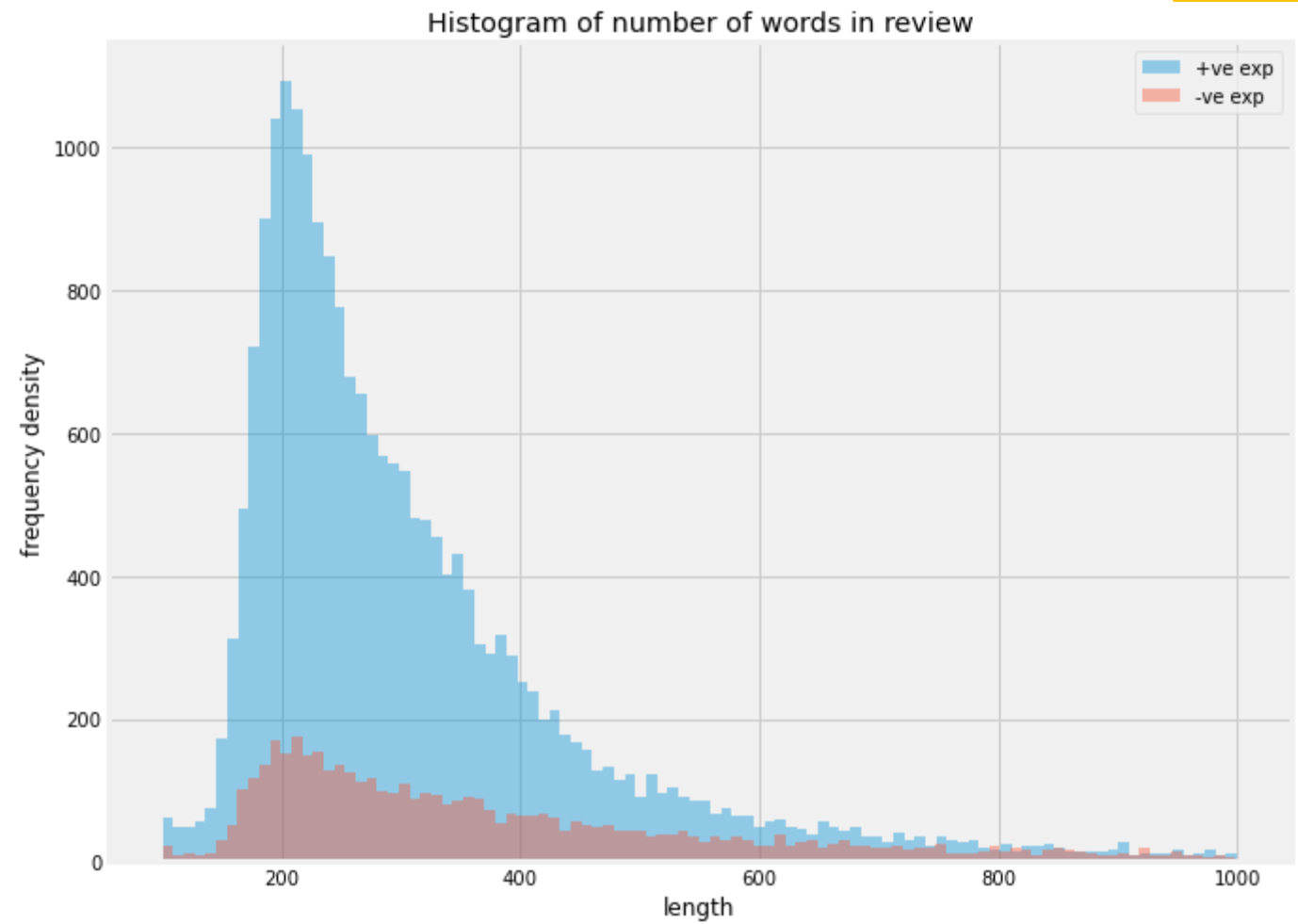
Helpful





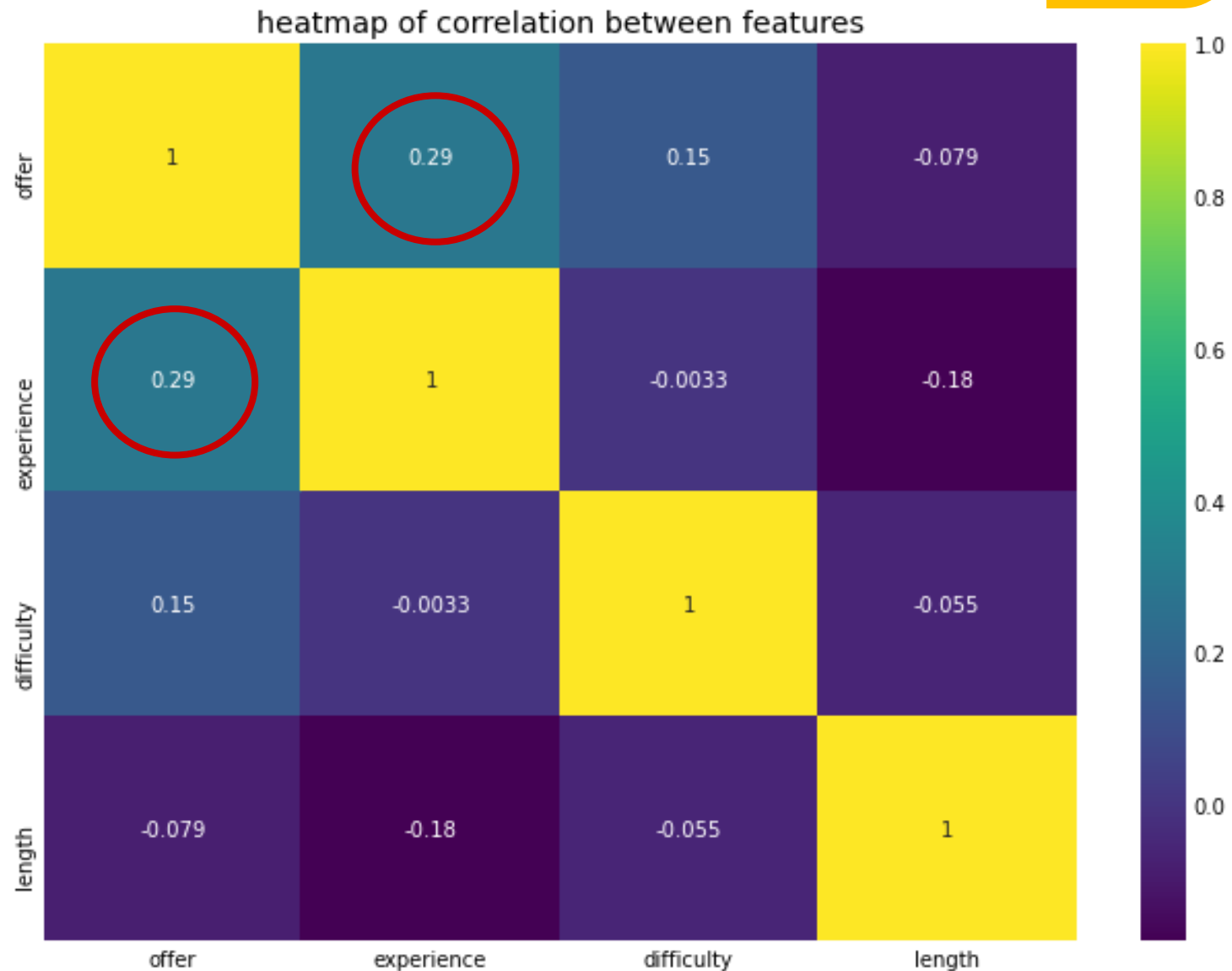
Count of words for positive and negative reviews

Distribution of word count for positive and negative reviews are generally the same



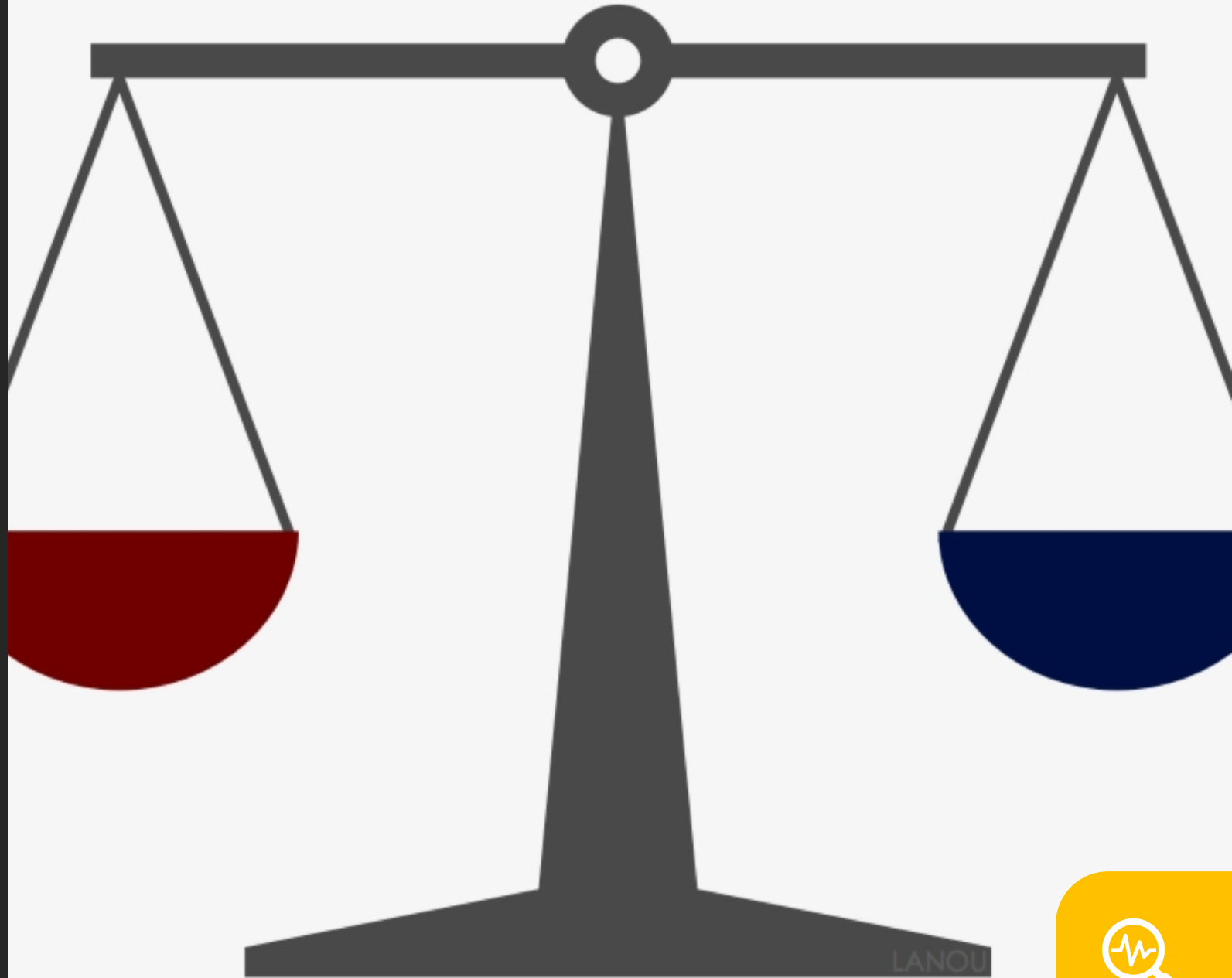
Correlation between different features and interview experience

The highest correlation
among the features and
interview experience is offer



Dealing with imbalanced class

- Scraped reviews from other companies on glassdoor and added the negative reviews into the original data.
- Increase negative reviews from 4,762 to 11,800
- There are 21,079 positive reviews for our majority class



Word pre-processing

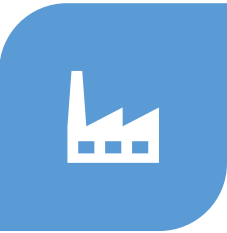
Remove flags, emojis and
html texts

Convert all words to lower
case

Lemmatizing each word,
remove stop words and
swear words

Removing words with less
than 2 characters





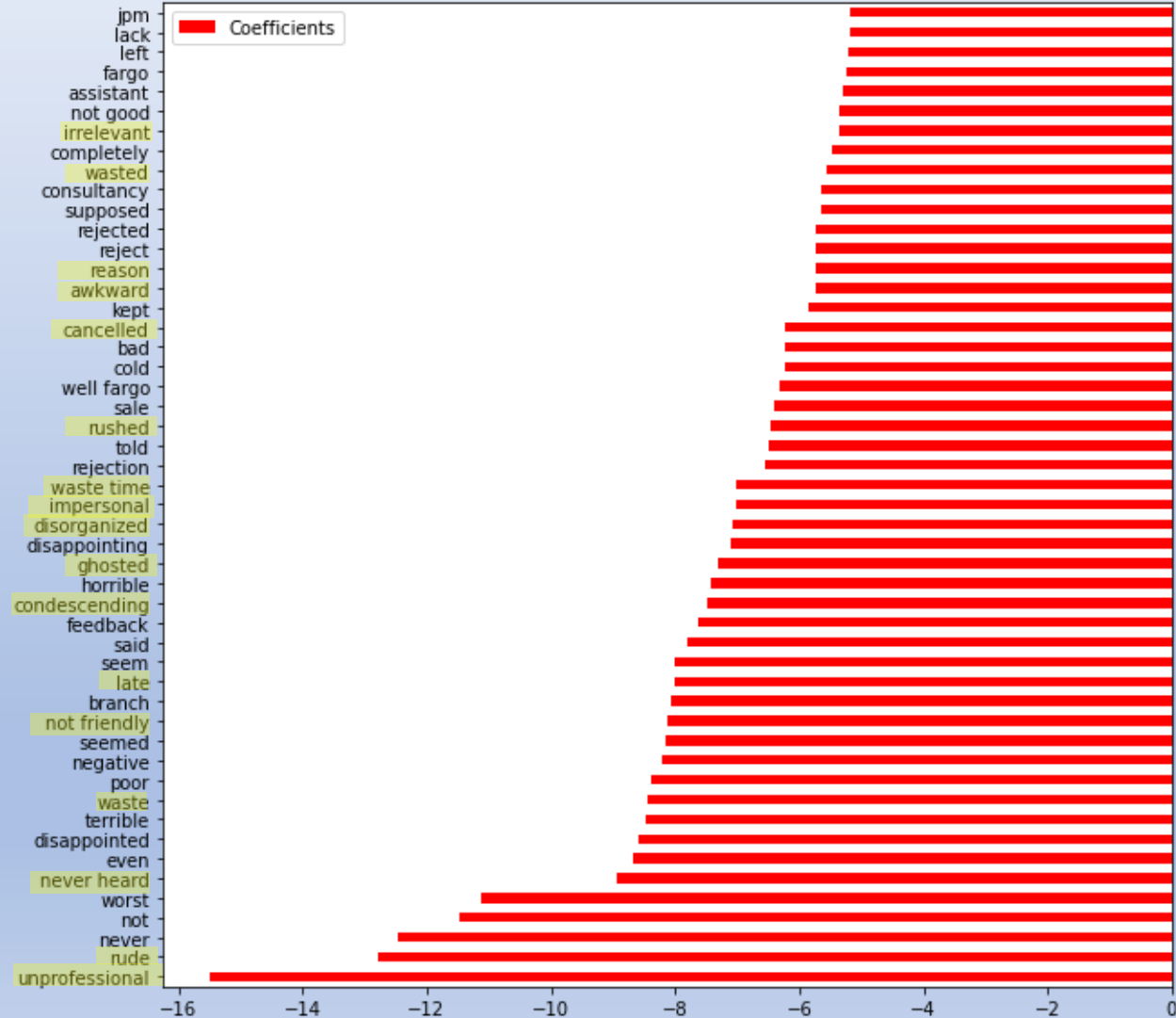
Model Selection and evaluation

| Model | Train Acc | Val Acc | Sensitivity | Specificity | ROC-AUC |
|-------------------------|-----------|---------|-------------|-------------|---------|
| K Nearest Neighbours | 0.7534 | 0.67609 | 0.9412 | 0.2025 | 0.6402 |
| Multinomial Naïve Bayes | 0.8243 | 0.8239 | 0.8833 | 0.7178 | 0.8791 |
| Logistic Regression | 0.8337 | 0.8349 | 0.9018 | 0.7152 | 0.8978 |
| Random Forest | 0.7803 | 0.785 | 0.8254 | 0.7127 | 0.8487 |
| XGBoost | 0.7298 | 0.7293 | 0.9957 | 0.2533 | 0.8637 |

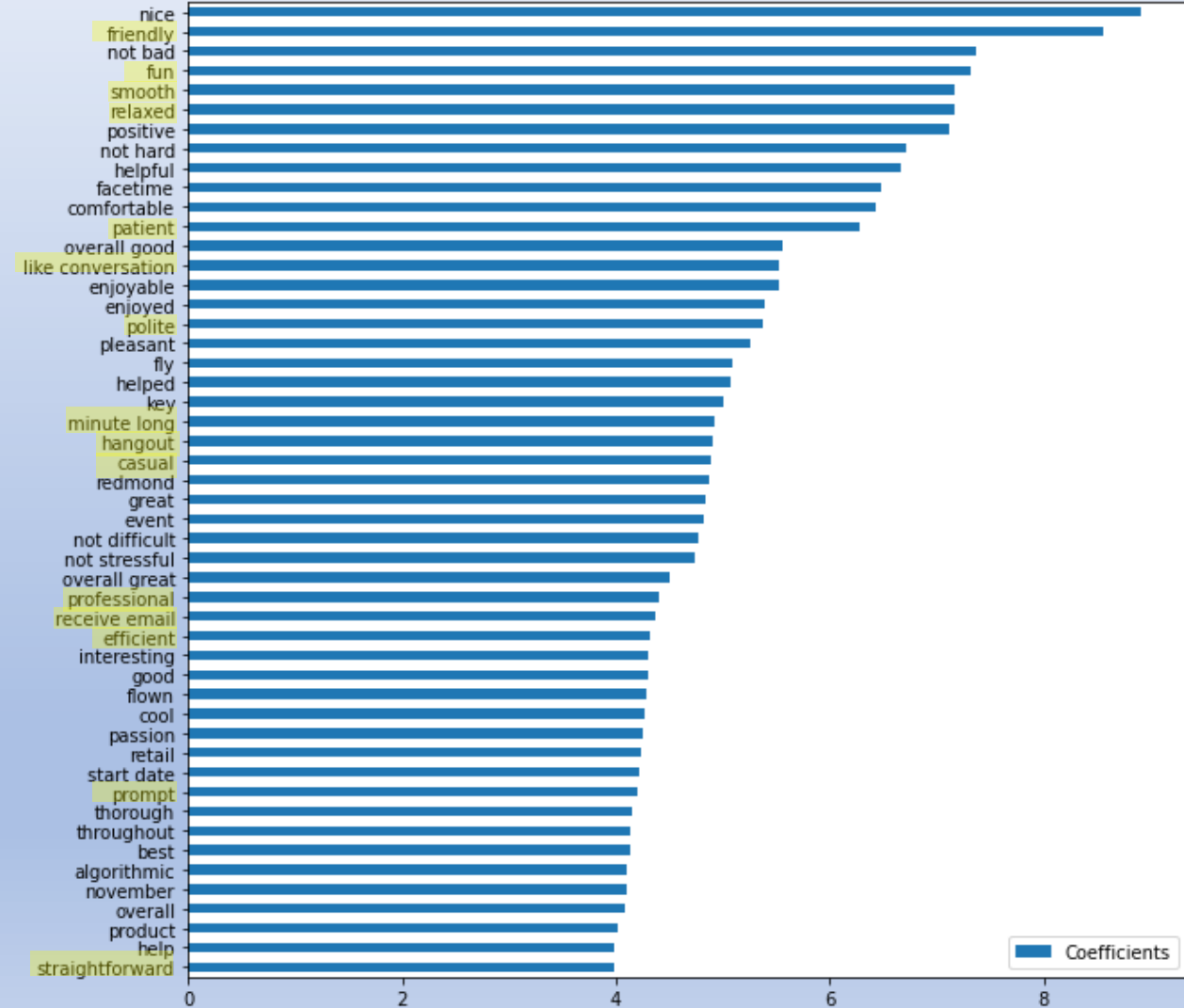
Model Findings



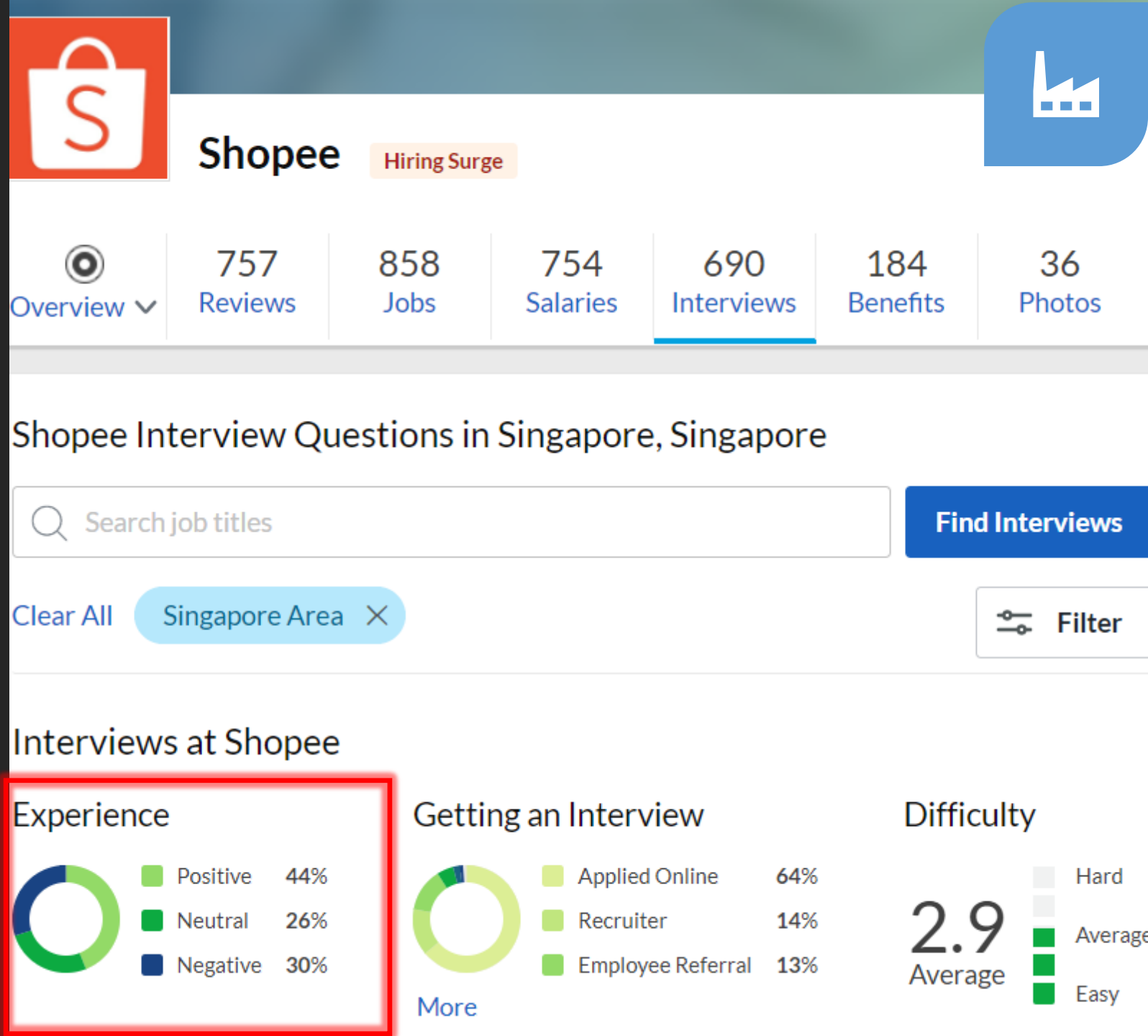
Words highly correlated to negative review



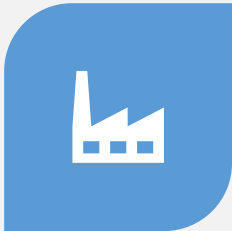
Words highly correlated to positive review



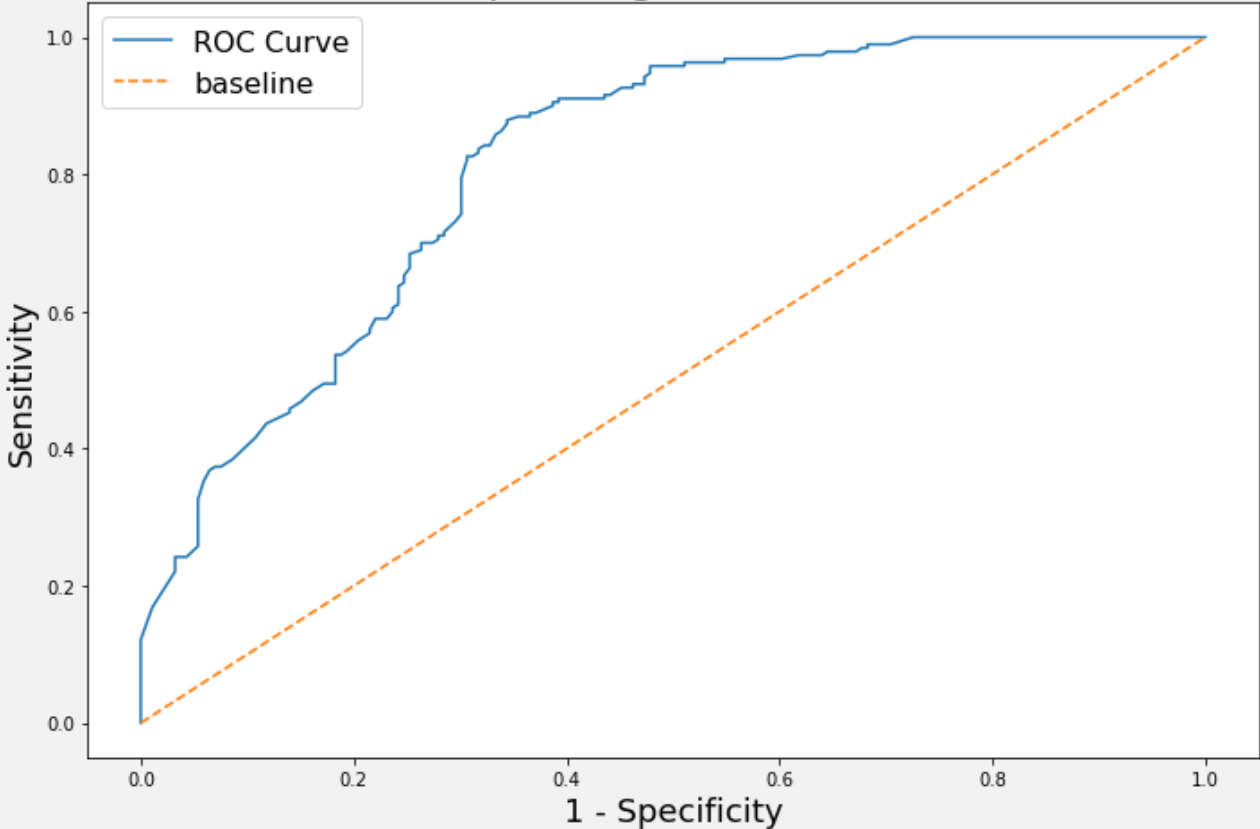
Production Model on Shopee data



Model Performance on real data



Receiver Operating Characteristic Curve



| Model | Accuracy | Sensitivity | Specificity | ROC-AUC |
|---------------------|----------|-------------|-------------|---------|
| Logistic Regression | 0.7606 | 0.8895 | 0.629 | 0.8143 |

| Confusion Matrix | | Predicted Columns | |
|------------------|------------------------|---------------------------|---------------------------|
| | | Predicted Negative Review | Predicted Positive Review |
| Actual Labels | Actual Negative review | 117 | 69 |
| | Actual Positive review | 21 | 169 |

Examples of Shopee interview reviews

Negative

One technical round. The interviewer provided a false link at the beginning. The interviewer does not even know Java and he kept asking how to design a dynamic size array in Java without reallocation. This is utterly disgusting.

The entire interview process was very disrespectful. They don't ask for your free time and simply give you a time. During the interview, they can be up to 30 minutes late! They don't seem to have much good things to say about the company too. A senior manager said work environment is very stressful.

Phone interview, sent me an email to schedule phone interview. Sent email afterwards to ask why I didn't get the role and if there's anything I can work on. Did not get a reply.



Positive

The recruiter is very nice and she will explain to you how the company views the most for this position. Phone interview is not difficult and then she directly gives me the chance for a 2nd round test interview. The test interview is about 1.5 hrs and the Python&SQL questions are very basic.

Phone interview with HR first then direct interview with the cat managers. both interviews were quite fast, within 30 minutes, i think the phone interview was shorter and the hr person was very friendly

HR Manager was friendly and provided information before the interview took place. There was a better understanding of the happenings and the overall reputation of the company at present. The questions were as expected (i.e. tell me more about yourself, why are you applying for Shopee, etc.

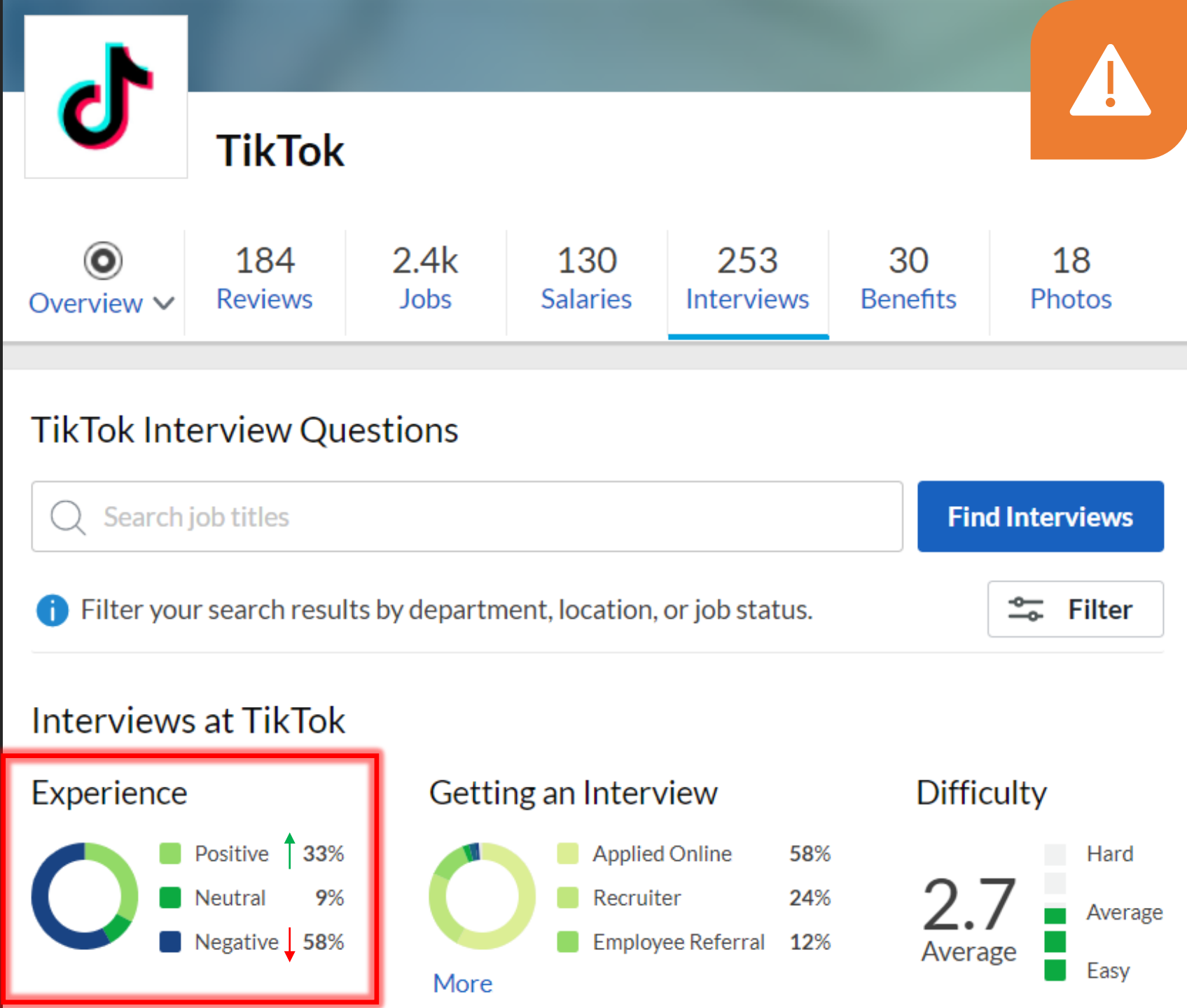


Limitations

- Words that are new to the corpus will not be considered when doing vectorizing transformation
- If the review is in another language, our model is not able to accurately predict the sentiment
- Sarcasm within a review might result in inaccurate prediction
- Logistic model assumes linear separability between different texts. However in reality, texts or comments are not exactly linearly separable
- Some words are subjective and scalable. Eg, nice is the word that has the highest coefficient for positive experience, however it is very subjective and there are too many variations of nice

Conclusion

- Companies that want to understand the sentiments of their reviews and gather insights can adopt this model
- Higher positive interview experience rating on their public profiles, achieve a better company brand and attract more talent





Questions