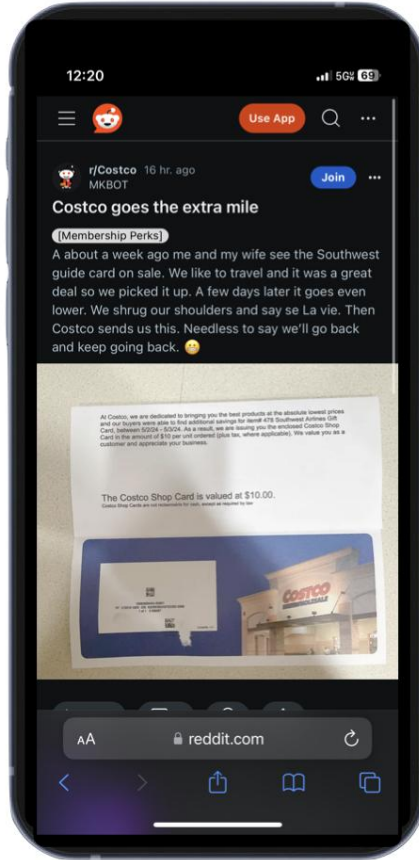




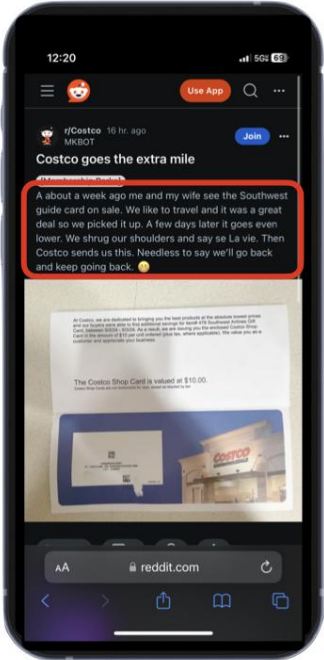
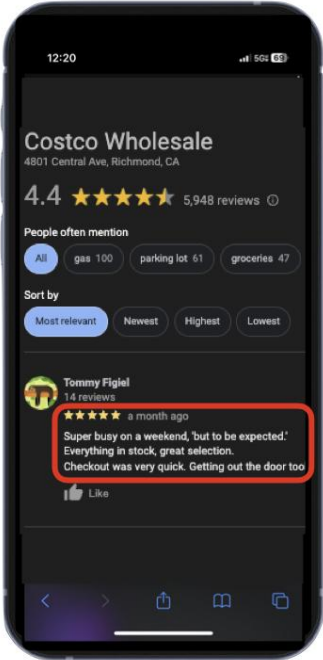
# How are people talking about my company online?

Vinicius Ambrosi, Gilyoung Cheong, Dohoon Kim, & Hannah Lloyd

How can you gather insights from online discourse to improve your products and company?



# Can google review comments accurately predict consumer ratings?



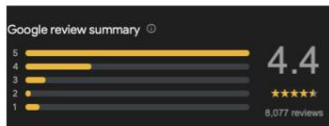
# Training models on imbalanced datasets of google reviews to accurately predict consumer ratings:

## Data Source & Collection

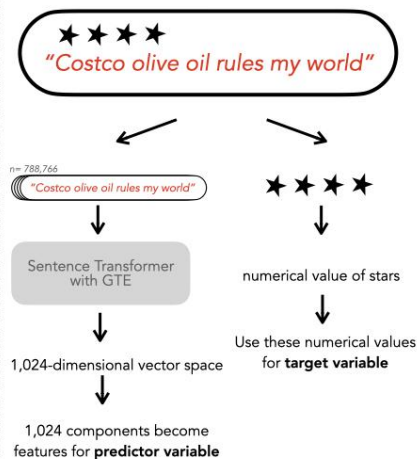
Target company:



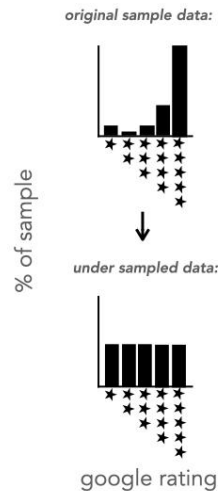
Scrape **788,766** Google Review comments and associated star rating for the target company.



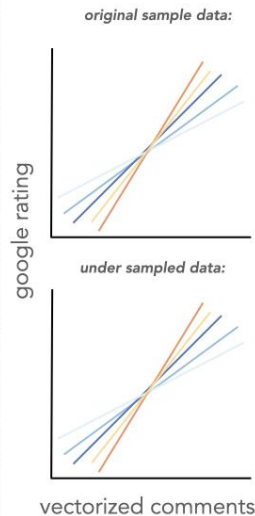
## Data Processing



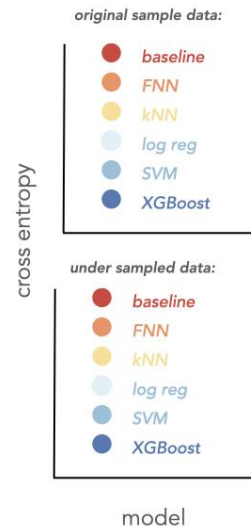
## Data Sampling



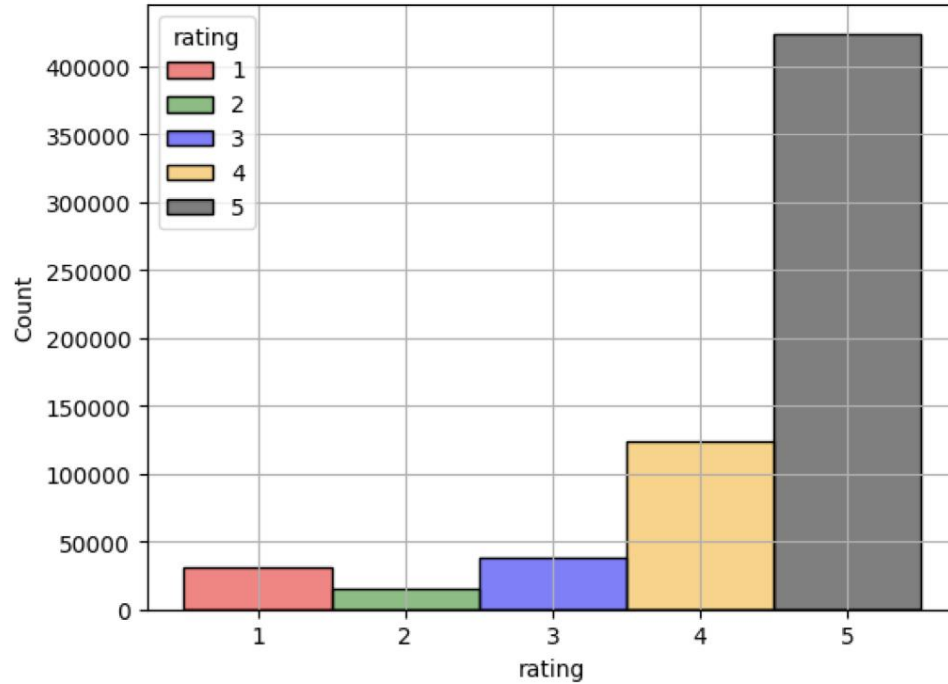
## Model Building



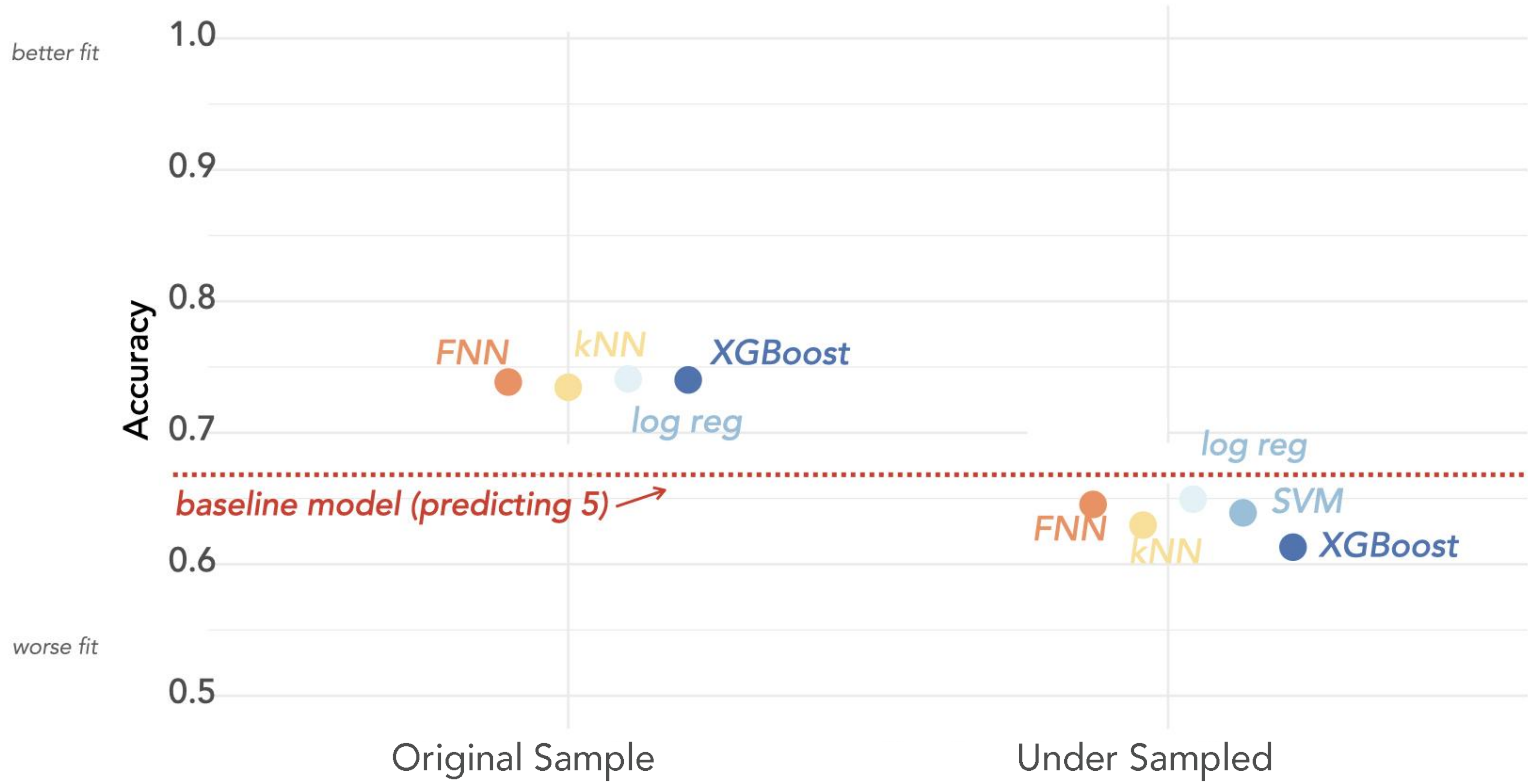
## Model Evaluation



# Skewed Distribution of Google Review Ratings

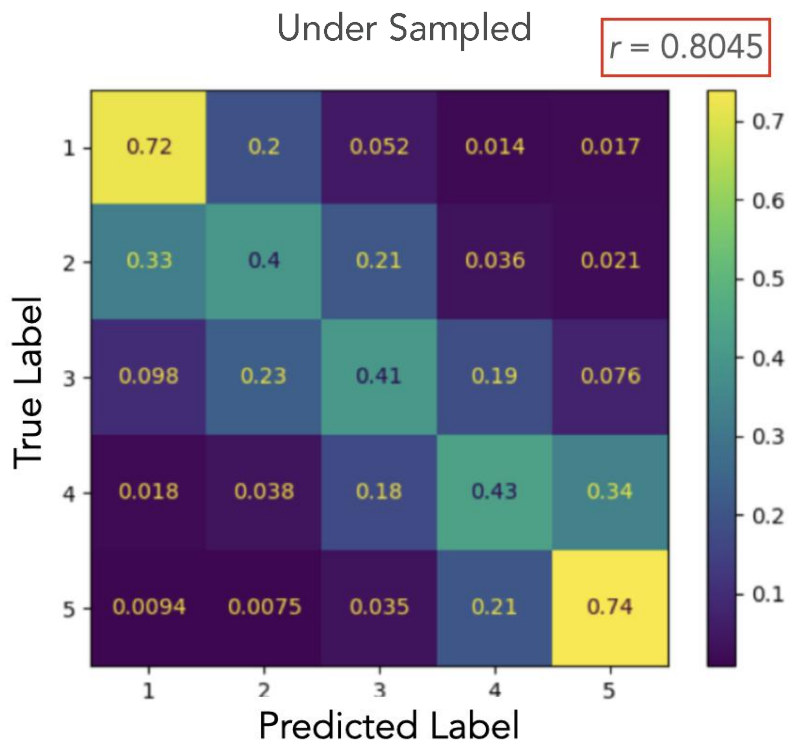
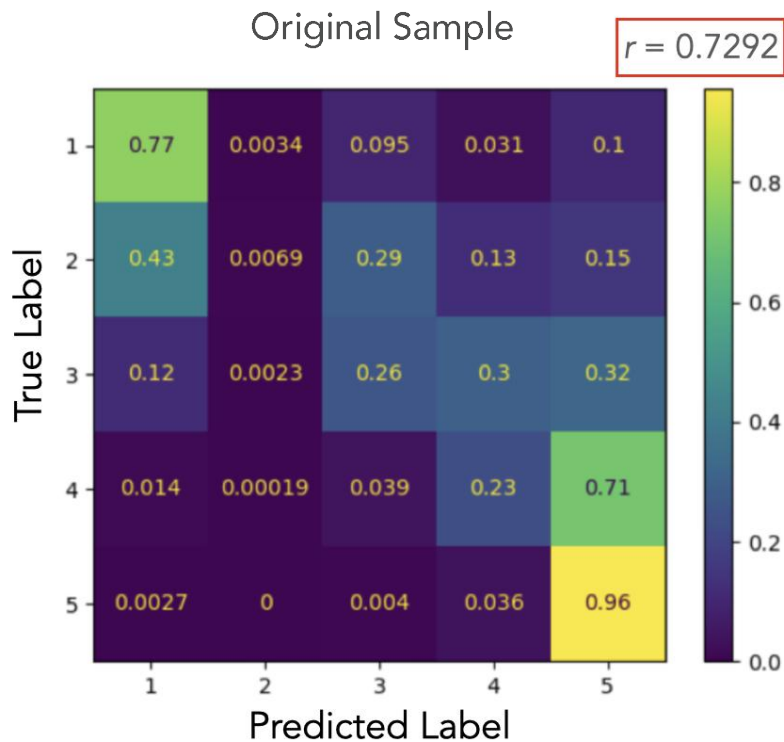


# Which Model Classifies Best? Comparing Accuracy Across Different Models

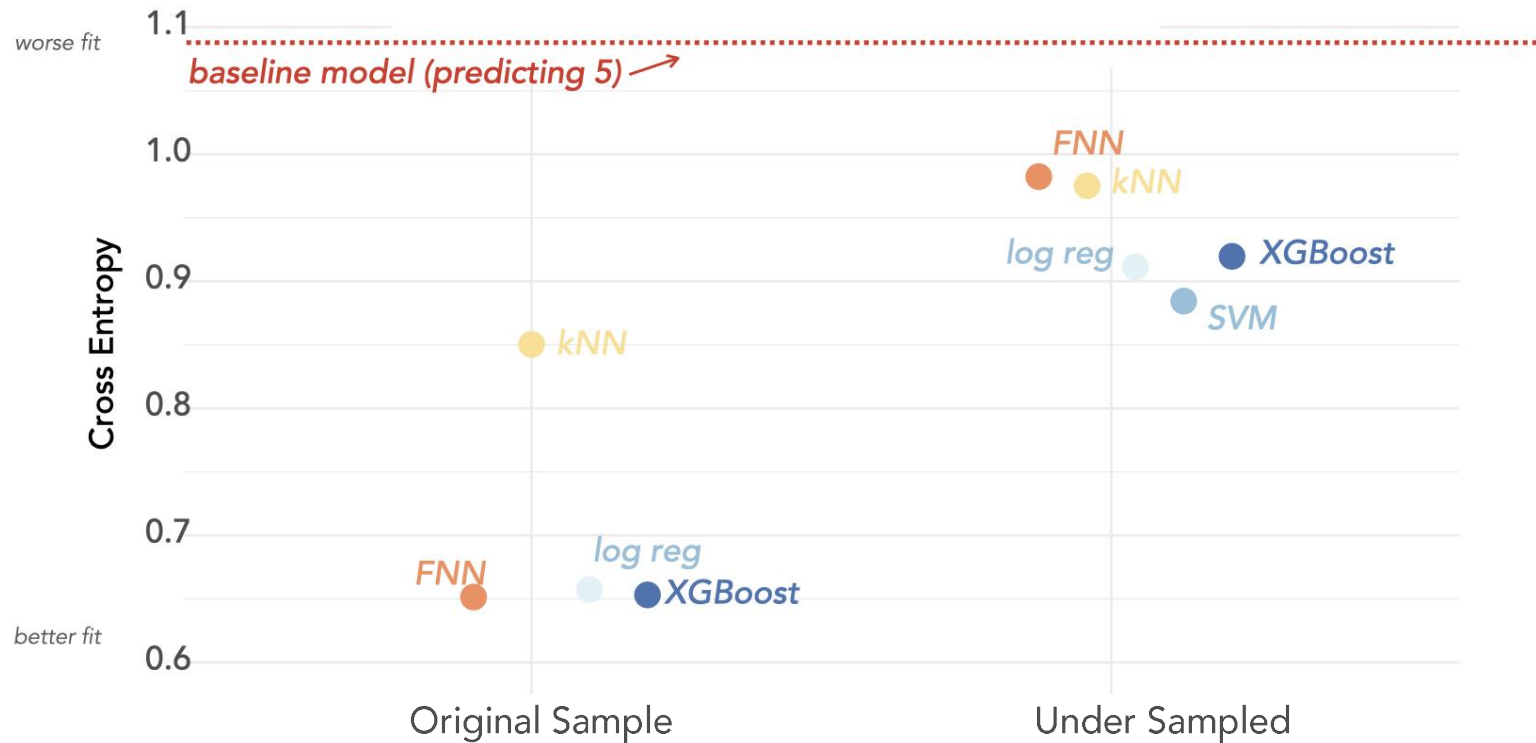


# Under sampled data classifies better when evaluating accuracy confusion matrices

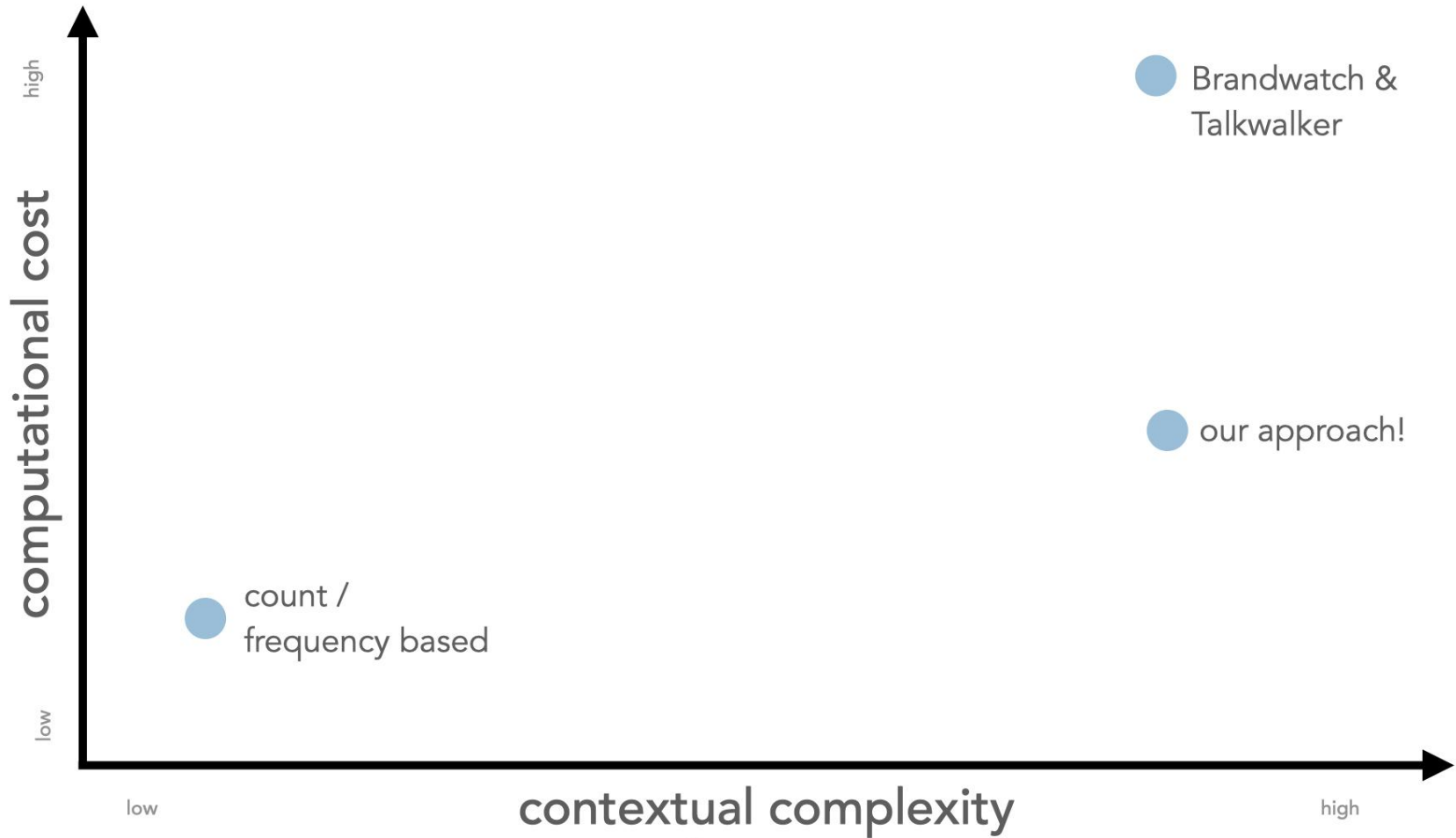
model:*logistic regression*



# Training models on imbalanced datasets of google reviews can predict consumer ratings:







# Practical Contributions

- Leveraged **under sampling** to train models that preserve contextual integrity and computational efficiency
  
- Discovered the importance of **evaluating model performance** beyond common metrics (i.e., accuracy)

# Practical Contributions & Future Directions

- Leveraged **under sampling** to train models that preserve contextual integrity and computational efficiency
  - **alternative dataset source**: test with data that has no ratings (i.e., reddit) & a different company
- Discovered the importance of **evaluating model performance** beyond common metrics (i.e., accuracy)
  - **alternative metrics**: accuracy doesn't tell the full story of the model performance



**Thank you!**