Yelp Pitch Deck

Team: Katrin Maliatski, Kaylyn Nguyen, Mahnoor Shahid,
Rimsa Shrestha, Jenna Woo, & Christina Zhu

Business Problem

Problem

Local businesses on Yelp struggle to interpret large volumes of unstructured customer reviews

Context

Although they have access to extensive user-generated data, small business owners often lack the tools needed to extract actionable insights

Business Impact

As a result, they miss valuable opportunities to improve products and enhance customer satisfaction, which ultimately harms customer retention, brand reputation, and revenue

6,990,280

Reviews

150,346

Businesses

11

Metropolitan areas

200,100

Pictures

Proposed Business Solution Approach

Objective: Enable local businesses to optimize their Yelp presence by using sentiment analysis to improve customer retention and operational decisions

Key Recommendations:

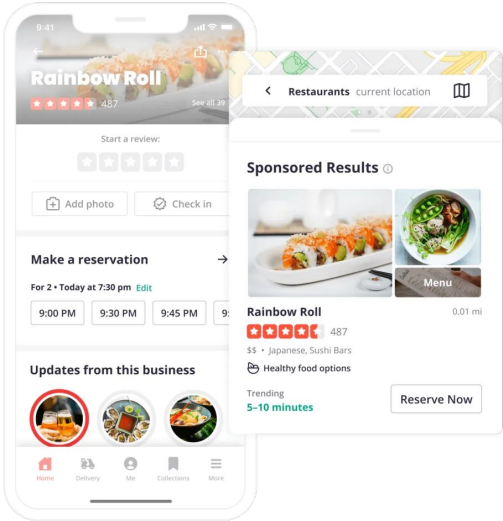
	Review Quality Optimization	Predictive Customer Targeting	Geographic Expansion Insights
Business Question	What keywords + tones drive higher ratings?	Who is most likely to revisit or engage again?	Where should we open a new location?
Action	Guide owners on using sentiment analysis on reviews	Use model predictions to segment and retarget high-likelihood users	Cluster review/check-in data to identify high-interest, low-competition areas
Outcome	Increase in star ratings by ~0.2–0.5 on average	Improved retention and re-engagement rates	15–25% performance boost via strategic location placement

Proposed Data Science Solution Approach

Goal: Use machine learning and NLP to perform sentiment analysis and deliver actionable insights for businesses on Yelp

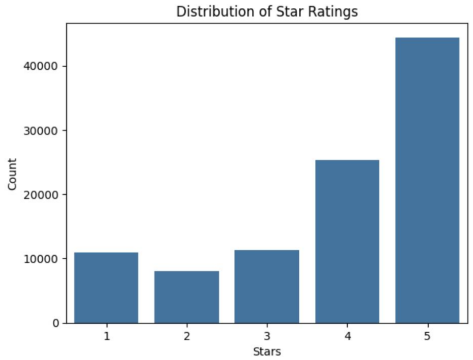
Approach:

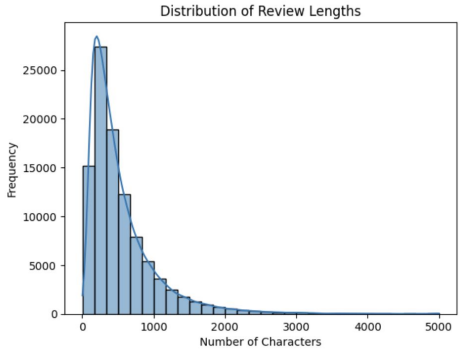
- **Problem:** Binary sentiment classification of customer reviews
- **Data:** Yelp Open Dataset (review text, star ratings converted to sentiment labels)
- **Users:** Local business owners, Yelp platform analysts
- **Target:** Binary sentiment classification of customer reviews
- **Features:** Text vectorization with Bag of Words, TF-IDF, and n-Grams
- **Algorithms:** Logistic Regression, Support Vector Machine, Random Forest
- **Evaluation Metrics:** Accuracy, Precision, Recall, F1 Score, ROC AUC
- **Deployment:** Model integration into customer insight dashboards

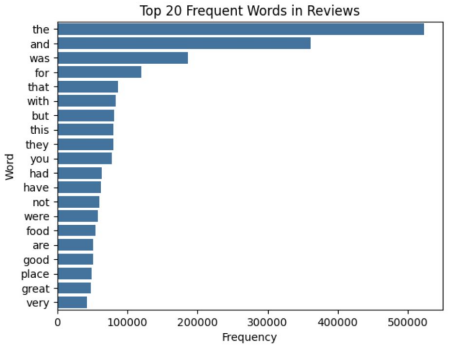


Data Science Process Model

Business Understanding: Yelp's role as a connector between users and local businesses
Data Understanding: Explore Yelp dataset to assess review patterns and user behavior
Data Collection: Gather review data with ratings
Data Preprocessing: Cleanse, label sentiment (positive/negative)
Feature Engineering: Vectorize text with three techniques (Bag of Words, TF-IDF, and n-Grams)
Modeling: Use Logistic Regression, Support Vector Machine, Random Forest to train and tune classifiers with GridSearchCV
Evaluation: Cross-validate model accuracy and business relevance of insights
Deployment & Monitoring: Set up real-time sentiment scoring







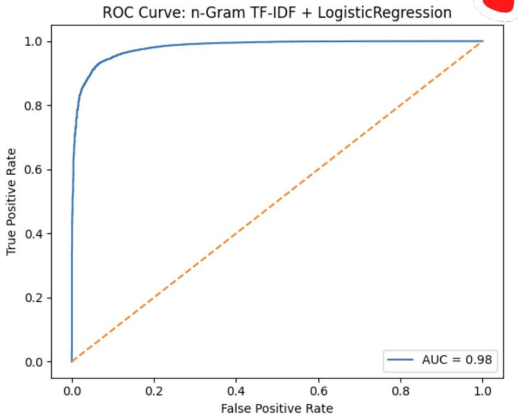
Sentiment Analysis Model

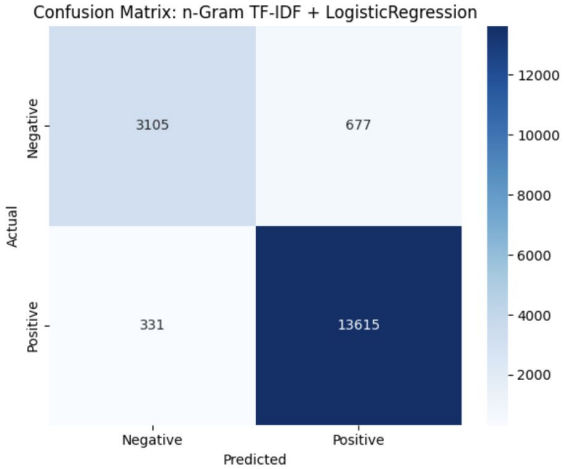
Text classification used to determine positive/negative sentiment to scale analysis of large text datasets, real-time insights

Top Performer of 9 Models: n-Gram + Logistic Regression

- Accuracy: 94%
- Precision: 94%
- Recall: 94%
- F1 Score: 94%
- ROC AUC: 98%

- Best combination of **accuracy, interpretability, and performance**
- Excellent at **minimizing false negatives**
  - Ensures few unhappy customers are missed
- Keeps **false positives low**
  - Avoids wasting resources on happy customers
- Ideal for early-stage sentiment monitoring and business decision-making





Data Science Process Model Implementation

Phase 1

Data collection and initial model development (Months 1-2)

Phase 2

Pilot testing with 50 coffee shops in Bay Area (Month 3)

Phase 3

Iteration & testing with 50 other small businesses in Bay Area (Month 4)

Phase 4

Dashboard development and API integration (Months 5-6)

Phase 5

Full deployment and monitoring system (Month 7)

Phase 6

Expand to 1000+ businesses across multiple categories (Months 8-12)

Expected Impact & Next Steps

Expected impact:

- Increase in average Yelp star ratings by guiding review response strategy
- 15-25% boost in local visibility through targeted geographic insights
- Improved customer retention by identifying and rewarding likely return users

Next steps:

- Finalize model selection based on performance evaluation
- Package recommendations into business-friendly deliverables (e.g., dashboards)
- Prepare demo use case for a sample business category (e.g., coffee shops)

Thank you!