DS 4002 - 2/12/2026

Predicting Restaurant Review Sentiment From Text

Group 15 - Group Leaders: Ciara Fasullo, Sahithi Palleti
Members: Anuti Shah, Ben Shults

Motivation & Question

Data & Methods

Modeling

Results & Implications

Motivation & Research Question

Why predict sentiment from review text?

- Online reviews influence consumer and business decisions
- Star ratings are easy, but **text explains why**
- Automating sentiment analysis allows businesses to analyze thousands of reviews efficiently

Can the language used in restaurant reviews reliably predict whether a customer rated their dining experience positively or negatively?

Hypothesis & Modeling Approach

The sentiment analysis model will predict with at least 70% accuracy:

⭐-⭐⭐: Negative review
⭐⭐⭐: Neutral review
⭐⭐⭐⭐-⭐⭐⭐⭐⭐: Positive review

Model:

- Supervised text classification
- Star ratings mapped to sentiment labels
- Baseline NLP models trained on text features

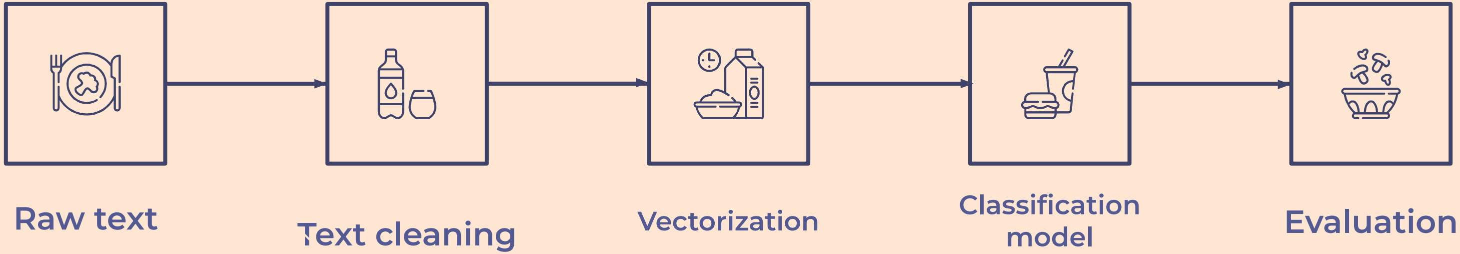
Data Overview

- Restaurant reviews with written text & star ratings
- Star ratings mapped to sentiment labels
- Reviews without written text removed
- Class distribution skewed toward positive sentiment

	Restaurant	Review	Rating
0	Beyond Flavours	The ambience was good, food was quite good . h...	5.0
1	Beyond Flavours	Ambience is too good for a pleasant evening. S...	5.0
2	Beyond Flavours	A must try.. great food great ambience. Thnx f...	5.0
3	Beyond Flavours	Soumen das and Arun was a great guy. Only beca...	5.0
4	Beyond Flavours	Food is good.we ordered Kodi drumsticks and ba...	5.0

Rating_num	Sentiment
5.0	positive
5.0	positive
5.0	positive
5.0	positive
5.0	positive

Analysis Pipeline

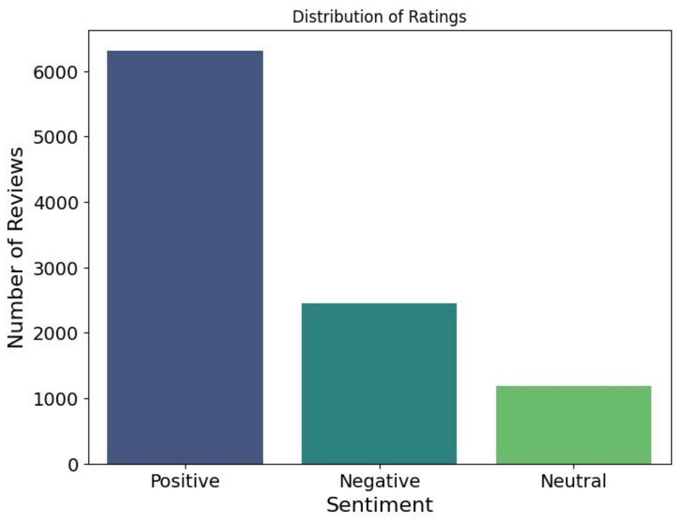


- Logistic Regression
- Naive Bayes
- Train/Test Split

Tricky Analysis Decision

Class Imbalance

- Majority of reviews are positive
- Accuracy alone can be misleading
- Reviews without written text removed
- Decision: use class-weighted models



Impact

- Prevents majority-class dominance
- Improves fairness across sentiment categories

Bias & Uncertainty

Potential biases

- Overrepresentation of positive experiences
- Self-selection bias in who leaves reviews

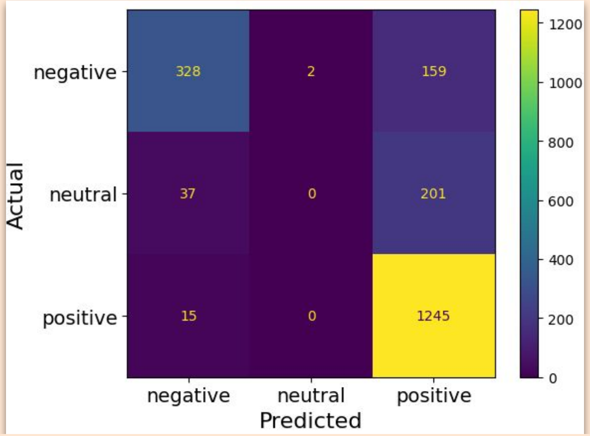
Sources of uncertainty

- Ambiguous language
- Neutral sentiment hardest to classify

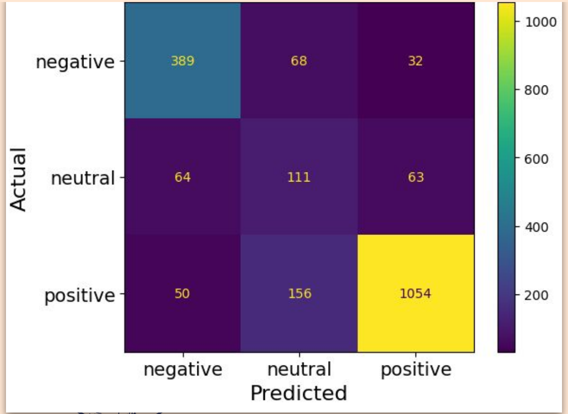
Validation

- Model comparison
- Weighted evaluation metrics

Naive Bayes Confusion Matrix



Logistic Regression Confusion Matrix



Results & Conclusions

Logistic regression

Best Model

78%

Accuracy

LR classification report

	precision	recall	f1-score	support
negative	0.77	0.80	0.78	489
neutral	0.33	0.47	0.39	238
positive	0.92	0.84	0.88	1260
accuracy			0.78	1987
macro avg	0.67	0.70	0.68	1987
weighted avg	0.81	0.78	0.79	1987

The model achieved 78% accuracy, exceeding our 70% benchmark. Therefore, our hypothesis is supported.

Next Steps

Incorporate bigrams and trigrams to capture word context

Improve neutral sentiment classification

Evaluate more advanced models

References

[1]“Restaurant Reviews,” *www.kaggle.com*. https://www.kaggle.com/datasets/joebeachcapital/restaurant-reviews

[2]B. Pang and L. Lee, “Opinion mining and sentiment analysis,” *Foundations and Trends in Information Retrieval*, vol. 2, no. 1–2, pp. 1–135, 2008, Available: https://www.cs.cornell.edu/home/llee/omsa/omsa.pdf

Github: https://github.com/benshults22/ds4002-project1-DataDawgs/tree/main

Questions?


4/24/2012 · Updated review
As much as I loved it, now I would not go back again.
Service was horrible, the food was trying too hard ...
No more ...
1 check-in here
5/6/2014
This place is a great lunch place... The food is great and
the waitress is so friendly! It's a little un
considering the area but I love it
regular place. The lunch speci
1/21/2017
The atmosphere is nice. I ordered steak and it was a small
portion compared to other competitors such as Kawa's
and Yamato. The prices were insane! I payed 26$ for a NY
strip when a Filet costs the same at Kawa's. My friend
ordered Sushi and the fish was RAW!!!! Never coming
back.