Yelp Dataset Sentiment Analysis

Literature Review Survey

Cole Hanrahan and Julia Sober
Text Mining & Analysis
Fall 2025

GEORGETOWN UNIVERSITY

Motivation – Why Sentiment Analysis on Yelp?

- Yelp reviews are full of **opinions**, but **star ratings** can be inconsistent or biased.
- Text analysis helps interpret **real customer sentiment** beyond stars.
- Bridges **AI** → **ML** → **NLP**: understanding language through computation.

Research Directions in Literature

- The field spans **three main research goals**:
  - **Binary Sentiment Classification** – positive vs. negative reviews.
  - **Star Rating Prediction** – predict 1–5 stars from text.
  - **Bias Analysis** – explore inconsistencies between text and rating.

Data & Preprocessing (Common Steps)

- Data: Mostly Yelp reviews, sometimes compared with Amazon, Twitter, IMDB.
- Common preprocessing steps:
  - Lowercasing, removing punctuation, stopwords.
  - Tokenization → focus on **adjectives** (key sentiment words).
  - **Negation handling:** "not good" → "not_good" (mixed results).
  - **Stemming:** Porter stemmer via NLTK.
- Split data: usually 70–30 or 80–20 train/test; some use k-fold validation.

Feature Engineering Approaches

- Different papers tested how best to represent text as numbers:
  - **Bag of Words (BoW)**: count word frequencies (still effective).
  - **Part-of-Speech Filtering:** use only adjectives.
  - **Sentiment Lexicons:** predefined word lists (e.g., Bing Liu).
- *Finding:* Raw text features often performed **as well as engineered ones** — overengineering sometimes adds noise.

Machine Learning Models Compared

Classical ML dominates:

• Ensemble methods (Bagging, AdaBoost) slightly improve results.

• Deep learning rarely used — traditional ML still performs well.

Model	Strength	Accuracy / RMSE
Naïve Bayes	Simple, fast	79%
Random Forest	Robust, handles noise	76%
Logistic Regression / SVM	Solid generalizers	50–55%
Linear Regression	Predicts stars	0.6

- Highlight specific papers here and show their different models

Evaluation and Findings

- **Metrics:** Accuracy, F1-Score, Precision/Recall, RMSE (for regression).
- **Trends observed:**
  - More features ≠ better performance (can cause overfitting).
  - Random Forest and Naïve Bayes remain top performers.
  - Regression models can estimate ratings fairly accurately.

Challenges and Gaps

- **Context & Sarcasm:** “Great job burning my pizza!” still reads as positive to models.
- **Lexicon mismatch:** Yelp slang & typos break standard sentiment dictionaries.
- **Imbalanced data:** Some ratings (2★, 3★) underrepresented.
- **Domain generalization:** Most models trained on restaurants only.

Interesting Findings

- **Simple ML still strong:** Naïve Bayes and Linear Regression remain competitive.
- **Feature design matters:** Focusing on adjectives helps; lexicons not always transferable.
- **Bias is real:** "Warm start" and subjective rating behaviors skew stars.
- **Future work:** Context-aware models (e.g., BERT) for sarcasm, multi-aspect sentiment, and business category generalization.

Conclusion and Next Steps

- Yelp sentiment research evolved from **binary classification → star prediction → bias correction.**
- Common thread: leveraging textual emotion for better business insights.
- **Our contribution:** continue this line by [state your project's focus — e.g., improving feature interpretability, using deep embeddings, or analyzing multi-aspect sentiments].