UNIVERSITY OF MICHIGAN

SCHOOL OF INFORMATION

# Natural Language Processing: Algorithms and People (SI 630)

## Project Proposal

### Product satisfaction and buyback intention prediction based on review analysis

Instructed by

Dr. David Jurgens

February 12, 2024



| Name | ID |
|---|---|
| Yanzhuo Cao | yanzhuo@umich.edu |

# Contents

# 1 Project Goals

The objective of this project is to develop a system capable of predicting customer ratings based on their textual reviews. This approach aims to reflect the nuances in customer shopping experience beyond simple binary sentiment outcomes, offering a multi-class output that ranges from 1 to 5. By analyzing tokenized words from reviews—focusing on those that are frequent yet not straightforwardly evaluative (e.g., not just ”good” or ”bad”)—the project seeks to uncover the aspects of products that customers care about most, such as price, quality, or features. Additionally, by incorporating whether a review comes from a verified purchase, the model aims to distinguish the credibility of feedback, further refining its predictive accuracy.

# 2 NLP Task Definition

This NLP task involves processing customer reviews to predict corresponding numerical ratings, effectively turning qualitative feedback into a quantitative assessment of customer’s shopping experience. The task will employ text analysis techniques to interpret the nuanced language of reviews. The outcome will be a model that not only predicts ratings but also identifies key factors influencing customer’s shopping experience by analyzing word frequency matrices for non-descriptive terms. Incorporating verification status as a model variable will allow for an evaluation of review credibility, enhancing the model’s utility in real-world applications.

# 3 Data

For this project, we will utilize the dataset available from the University of California, San Diego, specifically the ”Amazon Product Review Dataset” (Ni et al. (2019)) which includes customer reviews of products sold on Amazon. This rich dataset provides a comprehensive collection of review information, including:

- **’overall’:** The numerical rating given by the customer, ranging from 1 to 5.
- **’verified’:** Indicates whether the review comes from a verified purchase.
- **’reviewTimed’**: The ate when the review was posted.

- **'reviewerID', 'asin', 'reviewerName':** Identifiers for the reviewer, product, and the name of the reviewer.
- **'reviewText':** The textual content of the review.
- **'summary':** A short summary of the review.
- **'unixReviewTime':** The UNIX timestamp of the review time.

We plan to analyze a subset of this data, focusing on specific product categories to manage the dataset's size and complexity. Initial exploration indicates a diverse range of reviews, from brief comments to detailed feedback, offering a rich source for NLP tasks. The dataset includes millions of reviews, providing a solid basis for training and evaluating our models. For the scope of this proposal, we have examined a small sample that illustrates the dataset's structure and content, ensuring the text's legibility and the feasibility of our planned analyses.

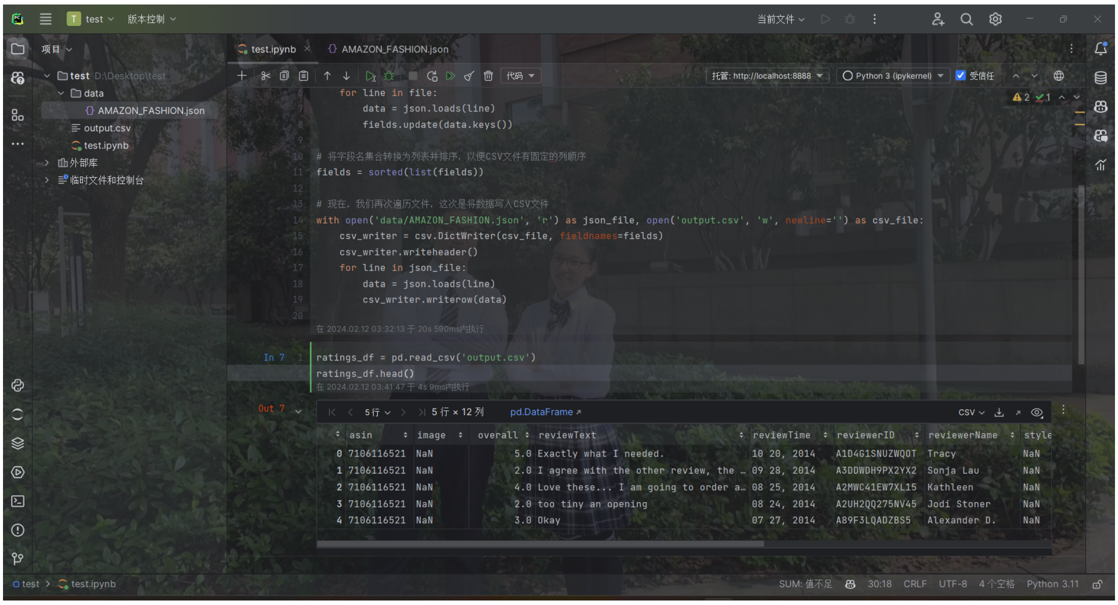


**Figure 1:** Data Example

# 4 Related Work

Several studies have explored customer satisfaction and sentiment analysis using NLP techniques, providing a foundation for our research. Alibasic and Popovic (2021) applied natural language processing to analyze customer satisfaction across airline reviews. They emphasized preprocessing steps like tokenization and stemming, and used sentiment analysis to differentiate between positive and negative sentiments. Their methodology showcases the importance of detailed text analysis in understanding customer feedback.

Another study focused on sentiment analysis on Amazon product reviews (Haseeb et al. (2023)). This research applied SVM and POS algorithms to categorize reviews and explored adverb extraction to refine sentiment analysis. Despite achieving high precision in some categories, the study highlighted challenges in accurately classifying neutral sentiments, underscoring the complexity of sentiment analysis in e-commerce.

Furthermore, text mining methods were employed for market intelligence and topic extraction, demonstrating the versatility of NLP applications in business contexts (Mohammad (2020)). These methods help in tracking and monitoring market trends, proving crucial for strategic business planning.

Our approach aims to extend these methodologies by predicting customer ratings from textual reviews more accurately. Unlike previous works that focused mainly on binary sentiment classification, our project seeks to provide a nuanced multi-class output that reflects a wider range of customer sentiments. Additionally, by analyzing non-evaluative terms and incorporating review verification status, our model intends to offer deeper insights into customer priorities and review credibility.

# 5 Evaluation

To quantitatively evaluate our NLP system, we will employ several metrics that are standard in classification tasks, including:

1. **Accuracy:** The proportion of total predictions that are correct. While informative, accuracy alone may not fully capture performance due to class imbalance.

2. **F1 Score:** A more comprehensive metric that considers both precision (the proportion of positive identifications that were actually correct) and recall (the proportion of actual positives that were identified correctly).

The F1 score provides a balance between precision and recall, making it particularly useful for our task where the distribution of ratings might be uneven.

3. **Confusion Matrix:** To visualize the performance across different rating categories, illustrating where the model performs well and where it confuses between classes.

For baselines:

1. **Random Performance Baseline:** As previously mentioned, predicting ratings randomly will establish a minimal expectation of performance.

2. **Most Frequent Class Baseline:** Predicting the most common rating in the training set serves as a naive yet instructive comparison, highlighting any significant improvement our model offers over simple heuristics.

Additionally, Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) will be used to measure the average error in predictions, providing insight into the prediction accuracy from another perspective.

These metrics together will offer a robust framework for evaluating the model's performance, ensuring a comprehensive understanding of its predictive capabilities and areas for improvement.

# 6 Work Plan

1. **Data Preparation (Weeks 1-2):** Begin with collecting and preprocessing the dataset. This includes cleaning the text, handling missing values, and preparing the data for analysis.

2. **Exploratory Data Analysis (Week 2):** Perform initial analysis to understand the dataset's characteristics, such as rating distribution, review length, and common words.

3. **Feature Engineering (Week 3):** Develop and extract relevant features from the text, including TF-IDF scores, sentiment scores, and other NLP-based features.

4. **Model Development (Weeks 4-5):** Experiment with various machine learning models (e.g., linear regression for rating prediction, classification models for sentiment analysis) to find the best performing model based on our evaluation metrics.

5. **Model Evaluation (Week 6)**: Evaluate the model using the defined metrics against the baselines. Adjust parameters and features based on performance.

6. **Analysis of Results (Week 7):** Analyze the model's performance, focusing on what features are most predictive of ratings and what the model reveals about customer sentiment and priorities.

7. **Finalizing and Documentation (Week 8):** Finalize the model based on the analysis, prepare a comprehensive report detailing the findings, methodologies, model performance, and insights on customer sentiment and satisfaction.

8. **Report Preparation (Week 9):** Prepare a report to summarize the project findings, methodologies, and implications for businesses.

This plan is subject to change based on emerging insights and challenges during the project execution phase.

# References


Alibasic, A. and Popovic, T. (2021). Applying natural language processing to analyze customer satisfaction.

Haseeb, A., Taseen, R., Sani, M., and Gul, Q. (2023). Sentiment analysis on amazon product reviews using text analysis and natural language processing methods.

Mohammad, A. (2020). Text mining on customer reviews for business intelligence. 40:260–269.

Ni, J., Li, J., and McAuley, J. (2019). Justifying recommendations using distantly-labeled reviews and fine-grained aspects. In *Proceedings of the Empirical Methods in Natural Language Processing (EMNLP).*